

Quality Control of Surface Wind Observations in Northeastern North America. Part II: Measurement Errors

ETOR E. LUCIO-ECEIZA AND J. FIDEL GONZÁLEZ-ROUCO

*Facultad de Ciencias Físicas, Universidad Complutense Madrid, and Instituto de Geociencias,
Universidad Complutense de Madrid–Consejo Superior de Investigaciones Científicas, Madrid, Spain*

JORGE NAVARRO

División de Energías Renovables, CIEMAT, Madrid, Spain

HUGO BELTRAMI

Climate and Atmospheric Sciences Institute, St. Francis Xavier University, Antigonish, Nova Scotia, Canada

JORGE CONTE

Global Forecasters S.L., Madrid, Spain

(Manuscript received 17 October 2016, in final form 20 September 2017)

ABSTRACT

A quality control (QC) process has been developed and applied to an observational database of surface wind speed and wind direction in northeastern North America. The database combines data from three datasets of different initial quality, including a total of 526 land stations and buoys distributed over the provinces of eastern Canada and five adjacent northeastern U.S. states. The data span from 1953 to 2010. The first part of the QC deals with data management issues and is developed in a companion paper. Part II, presented herein, is focused on the detection of measurement errors and deals with low-variability errors, like the occurrence of unrealistically long calms, and high-variability problems, like rapid changes in wind speed; some types of biases in wind speed and wind direction are also considered. About 0.5% (0.16%) of wind speed (wind direction) records have been flagged. Additionally, 15.87% (1.73%) of wind speed (wind direction) data have been corrected. The most pervasive error type in terms of affected sites and erased data corresponds to unrealistic low wind speeds (89% of sites affected with 0.35% records removed). The amount of detected and corrected/removed records in Part II (~9%) is approximately two orders of magnitude higher than that of Part I. Both management and measurement errors are shown to have a discernible impact on the statistics of the database.

1. Introduction

Performing meteorological measurements, data storage, and management is a delicate process that is never exempt of errors, despite the efforts and care invested in the task. For any meaningful use of these meteorological data, it is important to ensure, as much as possible, the validity of observations. The procedures used for this purpose constitute the so-called quality control (QC; e.g., Wade 1987; Gandin 1988; DeGaetano 1997; Shafer et al. 2000; Fiebrich et al. 2010; see section 1 in Lucio-Eceiza et al. 2017).

Some QC tests are focused on the detection of issues related to data transcription and collection or to errors that occurred during data manipulation, like the duplication of data sequences. Additionally, the standardization of practices that can vary across institutions like measurement units or reference times can be issues of importance for databases built using data from various source institutions. All these checks refer to *data management issues*. Additionally, there are other tests that address temporal or spatial consistency in the data that are designed to deal with errors often produced at the moment of sampling, as a result of instrumental malfunction, calibration or exposure problems. All these

Corresponding author: Etor E. Lucio-Eceiza, eelucio@fis.ucm.es

DOI: 10.1175/JTECH-D-16-0205.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

errors are generally of local nature and are less likely to depend on procedures established by the data source institution. We refer to these cases as *measurement errors*.

The present work summarizes the second part of a QC process applied to an historical data compilation of surface wind observations across northeastern North America (WNENA). [Lucio-Eceiza et al. \(2017, hereafter Part I\)](#) reports on data management issues, whereas the procedures described herein, Part II hereafter, are focused on the detection and removal/correction of measurement errors. [Part I](#) demonstrated that the problems related to data management had a very important impact on the surface wind data, with more than 90% of the data being modified during the process of unifying data transcription, collection, and storage, and ~0.1% of faulty records being deleted mainly due to intersite erroneous duplications of data sequences.

The goal of this paper is to analyze the problems related to measurement errors, to alleviate them with the help of data flagging/correction protocols, and to compare their extent with the data management issues detected in [Part I](#). Since the measurement errors are, by their very nature, independent of the dataset, the procedures presented herein are of universal applicability and thus easily translatable to other datasets. As with [Part I](#), for each test the behavior of the suspect records is addressed together with the statistics of occurrence in space, time, and data source. In both parts, an evaluation of the impact of errors on the statistics of the data is also provided. The final purpose of this work is to construct a surface wind speed and wind direction database of robust quality and wide spatial and temporal expanse that can be later used for the analysis of interesting phenomena specific to this region, like the analysis of extreme values ([Cheng 2014](#)), wind variability at different time scales and their relationship to large-scale modes of circulation (e.g., [Jiménez et al. 2008](#); [García-Bustamante et al. 2012](#)), high-resolution model validation (e.g., [Jiménez et al. 2010a](#)), or long-term electric production estimation (e.g., [García-Bustamante et al. 2013](#)) among others.

[Section 2](#) briefly describes the observational database. [Section 3](#) describes the methodologies of the QC process undertaken in this manuscript. [Section 4](#) provides an account of the results obtained at each phase of the QC procedure. The impact of the suppressed data is discussed in [section 5](#), and conclusions are given in [section 6](#). The purpose of [sections 5](#) and [6](#) is twofold, since they present the results specifically in reference to the treatment of measurement errors while offering at the same time a general view concerning the whole QC process in which the results obtained here are discussed in the perspective of those attained in [Part I](#).

2. Observational wind data

As it was extensively described in [Part I \(section 2\)](#), WNENA integrates the observations of 526 sites: 486 land stations distributed over eastern Canada (New Brunswick, Newfoundland and Labrador, Nova Scotia, Nunavut, Prince Edward Island, Ontario, and Quebec) and five northeastern U.S. states (Maine, Massachusetts, New Hampshire, New York, and Vermont), as well as 40 buoys distributed between the east coast of Canada and the Canadian Great Lakes. The area covers an approximate spatial extension of $2.2 \times 10^6 \text{ km}^2$. This database is the result of an aggregation of three different databases chosen for their availability and convenience, each one provided by a different institution: Environment Canada [EC; now known as Environment and Climate Change Canada (ECCC)], Fisheries and Oceans Canada Integrated Science Data Management division (DFO), and the operational global surface observations ([NCEP ADP OGSO 1980, 2004](#)) archived at the National Center of Atmospheric Research (NCAR). WNENA has an uneven distribution of stations, with higher spatial density over the southern area and along the coast, and lower density northward and inland. The database starts in 1953 and ends in 2010, spanning almost 60 years of hourly, 3-hourly, and 6-hourly recorded measurements. The initial quality of the data is disparate with the EC dataset having received some level of QC both in real-time mode and delayed mode ([MSC 2013](#)), and with the DFO and NCAR sites having received none to our knowledge [[Thomas and Swail \(2011\)](#), and ds461.0 and ds464.0 documentation pages]. In the compilation and development of WNENA, only simultaneously valid data pairs of both wind direction and speed are kept. The reader is referred to [Part I](#) for a thorough description of the datasets and instruments.

3. QC methodology

The QC that has been applied in WNENA is structured into six phases that deal with different issues (numbered in [Fig. 1](#)): 1) compilation; 2) duplication errors; 3) physical consistency in the ranges of recorded values; 4) temporal consistency, regarding abnormally high/low variability in the time series; 5) detection of long-term biases; and 6) removal of isolated records. The first three phases deal with data management issues and were addressed in [Part I](#). This manuscript focuses on issues that involve the last three phases in [Fig. 1](#). These can be regarded as measurement errors that are often related to instrumentation problems (temporal consistency and isolated records, phases 4 and 6, respectively), instrument calibration, and siting (bias detection, phase 5).

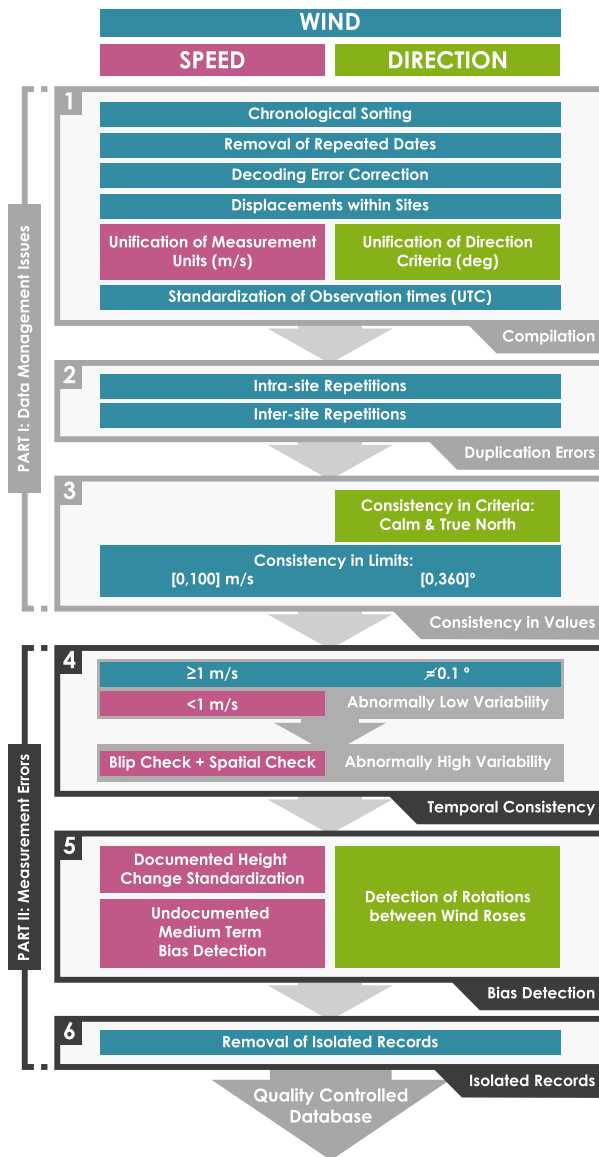


FIG. 1. Diagram describing the six phases of the QC process. Magenta (green) highlights checks that are applied only to wind speed (direction). Blue indicates tests applied to both variables. This paper deals only with measurement errors (last three phases); issues related to data management are treated in Part I.

The QC process follows a sequential structure designed to minimize potential overlapping between the various different phases. Some checks are common for both wind speed and direction, while others specifically address one of the variables. In Part I data identified as erroneous, such as duplicated chains of values, were removed. However, in Part II, erroneous values are just flagged for posterior removal (FR) or correction (FC). During the process, the FR data are temporally removed in order to establish the thresholds of subsequent steps

but are reverted at the time of applying them so that each record can be flagged by more than one step. The FC data, however, are kept corrected permanently as, for instance, in the case of documented height changes. All the flags are stored in a separate track file that codifies each step in a unique way for easy identification and eventual reversal. This section makes a methodological description of each phase, while the presentation of results and the illustration of specific cases will be addressed in the next section. A summary of the procedures is collected in Tables 1, 2.

a. Phase 4: Temporal consistency

These checks (phase 4 in Fig. 1) analyze the consistency of the temporal variability of the wind series. They target two different kinds of extreme behavior within the time series: periods with abnormally low or abnormally high variability.

1) ABNORMALLY LOW VARIABILITY

Periods with an inordinately small variability in wind speed and direction are typically the result of damaged instruments, caused by dust, corrosion or icing conditions, or as result of faulty communications between an instrument and the datalogger (Shafer et al. 2000). Various approaches have been taken to identify such errors. Some studies look for low-variability periods at relatively long time scales (e.g., 24 h in Shafer et al. 2000; one month in Hubbard et al. 2005). They compare the standard deviation of data in a given predefined moving window with a previously established threshold value. Periods with standard deviation values below this limit are flagged. Other studies search for constant data sequences (i.e., zero variance) of suspicious length at shorter time scales (e.g., minutes in Jiménez et al. 2010b; hourly in Meek and Hatfield 1994; 3-hourly in DeGaetano 1997). The unrealistically long constant value chains can be identified by establishing a threshold length. Maximum threshold lengths can be either arbitrarily imposed (Meek and Hatfield 1994; Durre et al. 2010; Dunn et al. 2016) or estimated from the sample statistics (Jiménez et al. 2010b). Alternatively, the faulty constant sequences can be identified with the help of an auxiliary variable (e.g., pressure; DeGaetano 1997).

The approach presented herein targets the search of constant data sequences. Direction sequences corresponding to records of 0° are excluded from the analysis, since this value was imposed for 0 m s^{-1} wind speed situations (Part I, section 3c). For wind speed a distinction is made between constant periods larger than or equal to 1 m s^{-1} and low wind speeds ($< 1 \text{ m s}^{-1}$), also regarded as calms in a loose sense (Jiménez et al. 2010b; MSC 2013). Measurements at low wind speeds ($< 1 \text{ m s}^{-1}$) are more

TABLE 1. Summary of the procedures carried out in Part II related to abnormally low and high variability (phase 4). The meaning of the abbreviations/symbols is at the bottom of Table 2. For a more detailed explanation, refer to section 3a.

Test	Procedure	Comments
Phase 4: Abnormally low variability (constant sequences/periods)		
Periods with excessive MISS ($WS \geq 1\text{ms}^{-1}$, $< 1\text{ms}^{-1}$; $WD \neq 0^\circ$)	Count MISS% in three periods: $[P]_{\pm 24\text{ h}}$: If any two periods have MISS% $> 90\% \Rightarrow [P] \bullet$	All $[P]$ segregated: $360/36/16/8$ vane points (WD), $0.1/0.3/0.5/1\text{ m s}^{-1}$ bins (WS) and $1\text{ h}/3\text{ h}/6\text{ h}$ (resol)
Longest valid period ($WS \geq 1\text{ms}^{-1}$) ($WS < 1\text{ms}^{-1}$) ($WD \neq 0^\circ$)	$[P]_{\text{Length}} \geq P_{95} + 8\text{IQR} \Rightarrow [P] \bullet$ $[P]_{\text{Length}} \geq P_{95} + 7.5\text{IQR} \Rightarrow [P] \bullet$ $[P]_{\text{Length}} \geq P_{95} + 15\text{IQR} \Rightarrow [P] \bullet$	All $[P]$ segregated as above
Spatial consistency ($WS < 1\text{ms}^{-1}$)	a) Construct a $[ra]$ for $[P] \pm 24\text{ h}$ b) Analyze $[ra]$ during $[P]$ (in) $\pm 24\text{ h}$ (out): 1. $[ra]$ declines in speed during $[P] \Rightarrow [P]$ real 2. $[ra]$ does not decline $\Rightarrow [P] \bullet$ 3. $[ra]$ partially declines \Rightarrow partially $[P] \bullet$	WD \Rightarrow unflag if $[P]$ is recorded in preferred direction $[ra]$: average of the five closest sites to target The analysis is done record by record $[ra(t)]$: $R_{[ra(t)]} = \frac{ra(t)_{\text{in}} - ra_{\text{min}}}{ra_{\text{max}} - ra_{\text{min}}} < 0.33; \forall [ra(t)]_{\text{(in)}}$ $R_{[ra(t)]} \geq 0.33; \forall [ra(t)]_{\text{(in)}}$ $[P(t)] \bullet$ for t when $R_{[ra(t)]} \geq 0.33$
Phase 4: Abnormally high variability (only for WS)		
Temporal check	For each $v(t)$, analyses Δv between closest Δt : $ \Delta v > P_{95} + 8.9\text{IQR} \Rightarrow \Delta v^*$	Analysis includes $t \pm [1, \dots, 23]\text{ h}$ Spike: two consecutive ($\Delta v > 0; \Delta v < 0$) $^* \Rightarrow v \bullet$ Dip: two consecutive ($\Delta v < 0; \Delta v > 0$) $^* \Rightarrow v \bullet$
Spatial check	Compares each $v(t)$ with neighbors: $v_{\text{thres}}: P_{95} + 8.9\text{IQR}$; if $v(t) > v_{\text{thres}} \Rightarrow v^*$	Neighbors: $\Delta r \leq 300\text{ km}$, $\Delta z \leq 500\text{ m}$ Step: ($\Delta v > 0$ and $v(t) > v_{\text{thres}}$) $^* \Rightarrow \Delta v \bullet$ Long period: $\forall v(t) > v_{\text{thres}}$ between two steps $\Rightarrow v \bullet$

TABLE 2. Summary of the procedures carried out in Part II related to bias detection and isolated values (phases 5 and 6, respectively). The meaning of the abbreviations/symbols is at the bottom of this table. For a more detailed explanation, please refer to [section 3b](#) and [3c](#).

Test	Procedure	Comments
Phase 5: Bias detection		
Documented height changes (WS)	$v_{z_{\text{ref}}} = v_z \frac{\ln(z_{\text{ref}}/z_0)}{\ln(z/z_0)}$	z_{ref} is the last documented height. z_0 derived from USGS EROS GLCC Data Base
Undocumented biases (WS)	If $\overline{v}(t) \notin (0.5, P_{95} + 3.7\text{IQR}) \Rightarrow \overline{v}(t)^*$ If $\sigma_v(t) \notin (0.044, P_{95} + 7.5\text{IQR}) \Rightarrow \sigma_v(t)^*$ If $\text{cv}(t) \notin (0.22, P_{95} + 7.5\text{IQR}) \Rightarrow \text{cv}^*$	$[\overline{v}(t), \sigma_v(t), \text{cv}(t)]$: deseasonalized, daily, 15- d_{avg} 2 weeks of only 1 of $[\overline{v}(t), \sigma_v(t) \text{ or } \text{cv}(t)]^* \Rightarrow [P] \bullet$ 1 week of 2 of $[\overline{v}(t), \sigma_v(t) \text{ or } \text{cv}(t)]^* \Rightarrow [P] \bullet$
Vane rotations (WD)	Evaluate RMSE between wind roses on consecutive $[P]$: 1. Rotate the roses with respect to the other 2. The angle of minimum RMSE gives the shift Correct: add/subtract shift \bullet to match most recent rose	The procedure is run twice. First time: $[P]$: one natural year \Rightarrow shift* second time: $[P]$: period between shift* \Rightarrow shift \bullet Minimum detected angle: 20°
Phase 6: Isolated records		
Isolated values (WS and WD)	if $[P] < 24$ h and surrounding 48 h = MISS $\Rightarrow [P] \bullet$	
\forall	Key of abbreviations/symbols	
Every value/record	Suspect record/period	\bullet Flagged record/period for removal/correction
cv	Δr	Distance between two sites
Δv	Δz	Elevation difference between two sites
IQR	MISS	Missing record(s)
$[P]$	$[ra]$	Regional average
RMSE	$\sigma_v(t)$	WS standard deviation
$v, v(t)$	$\overline{v}(t)$	Mean WS
WS	15- d_{avg}	15-day centered moving average

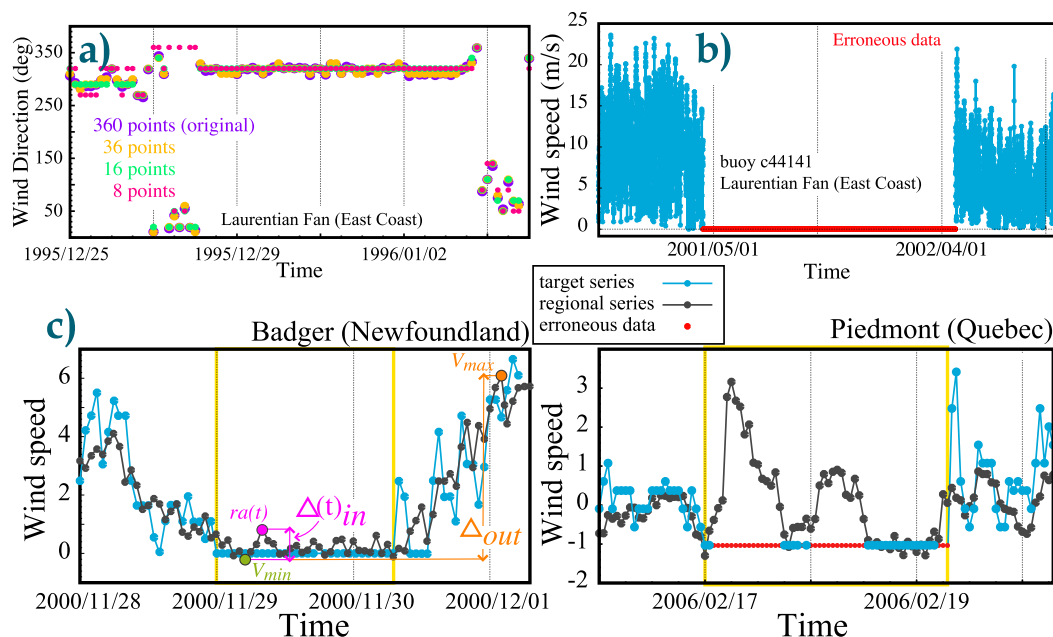


FIG. 2. (a) Example of the effect of precision degradation for the detection of constant wind direction periods for buoy c44141 located at the Laurentian Fan (east coast; DFO). The original precision (purple) has been successively degraded to lower precisions typical in the database (other colors). The largest constant period at the lowest precision occurring between 28 Dec 1995 and 4 Jan 1996 (159 records) was nonexistent at the highest precision. (b) Example of an erroneous calm episode that expands over a year at buoy c44141 (Laurentian Fan, highlighted in red). (c) Examples for calm detection at sites 8400301 (Badger, Newfoundland; EC; left) and 7026042 (Piedmont, Quebec; EC; right). The target series (light blue) and the regional series (black) are indicated. The analyzed calm periods are highlighted (yellow bars). The elements for the calculation of Eq. (1) (pink), V_{\min} (green), and V_{\max} (orange)—are shown: the individual records considered erroneous (red) are indicated. Both the target series and the regional series are standardized and are thus dimensionless. The regional series have been rescaled to the test's variance for visualization purposes.

prone to be affected by the deterioration of anemometers than any other measurements (WMO 2008), which can raise the anemometer initial wind speed response or artificially increase the length of calm events. Institutional efforts to improve the accuracy of low wind speeds (typically $\leq 2 \text{ m s}^{-1}$) via indirect Beaufort scale estimation have not always been systematically applied at sites, which can also lead to representativeness problems of low values through time and when comparing neighboring sites (DeGaetano 1998). The recent transition to ultrasonic anemometers at some sites can alleviate these problems (see Part I, section 2), but overall suspiciously long calm situations are more frequent and persistent than repetitions at higher wind speeds and are here grouped separately for the analysis.

The following methodology consists of three tests: the first two tests address wind direction and both high and low wind speeds, while the third test targets only low wind speeds/calm periods. Prior to their application, the sequences are classified in 12 different resolution \times precision (RP) groups that will be evaluated separately: hourly/3 hourly/6 hourly for resolution, and 360/36/16/8

points of the compass for direction and 0.1/0.3/0.5/1 m s^{-1} for speed. This is a highly recommended practice, since these differences can artificially affect the number of detected periods and their duration, as we have verified. See the example in Fig. 2a, where data precision has been artificially degraded, leading to the occurrence of longer chains of constant values. The minimum length of repeated values considered for evaluation is five records.

The first test searches for constant sequences surrounded by a large proportion of missing data, which can be indicative of operational problems. The test evaluates the percentage of missing data during each constant sequence and during the preceding and following 24-h intervals. The sequence is flagged as erroneous if any two of these three percentages exceeds a given threshold. The threshold was set at 90%, as it was observed that lower values tended to erroneously flag correct sequences of sites limited to daylight measurements (see Part I, section 2).

The second test evaluates the statistical likelihood of a constant data sequence depending on its length. The suspicious periods can span from several hours to

months long, such as the example shown in Fig. 2b for an erroneous calm that lasted around a year. Prior to the evaluation, the constant sequences are segregated by site \times resolution \times precision (SRP) in 12 RP groups for each site of the database (i.e., 526×12) and sorted by their length. For each distribution a nonparametric threshold is established, based on the distance given by $P_{95} + n\text{IQR}$, where P_x stands for the x th percentile and $\text{IQR} = P_{75} - P_{25}$ is the interquartile range of the distribution. The periods exceeding these thresholds are flagged as erroneous. The parameter n has been heuristically obtained in order to find a balance between the number of flagged cases and the number of false positives. The latter threshold, which erroneously flagged valid observations, is kept at a rate of $\sim 20\%$ (Durre et al. 2010), a practice followed for all the steps in this work. Thus, n is 15 for wind direction, 8 for noncalm speeds, and 7.5 for calm periods. For sites with RP distributions with fewer than 100 cases, thresholds are obtained using the sequences of all the sites. Additionally, constant wind direction sequences are more likely for the preferred directions at a site, particularly for low precisions. These sequences have been considered valid despite exceeding the threshold.

A third test is applied only to constant sequences at low wind speeds. It is based on the spatial consistency of the wind variability at a site and their neighbors and is able to detect periods that, albeit being erroneous, were overlooked by the previous test due to their short length. In real calm situations, well-chosen neighbor stations should also experience a decline in wind speed. A regional reference is constructed by selecting the five closest and best correlated, $r > 0.4$, sites during the 30 days centered at the time of each calm period. The calm periods are excluded to calculate these correlations. From each selected site, the period spanning 24 h before and after the constant data sequence is selected and standardized to 0 mean and unit standard deviation. Finally, a regional average, $ra(t)$, is constructed for each time step t . If the values of $ra(t)$ during the calm situations at the test site drop below a certain level, this is considered an indication that the wind was low and that the zero values at the test site were plausible and are not flagged.

The evaluation of the behavior of $ra(t)$ is done by considering the range of wind speeds during the supposed calm relative to that immediately before and after. Therefore, the minimum wind speed value (V_{\min}) during the candidate calm and the maximum wind speed value (V_{\max}) immediately 24 h before and after the calm are considered. The ratio at each time step,

$$R_{ra}(t) = \frac{ra(t) - V_{\min}}{V_{\max} - V_{\min}} \equiv \frac{\Delta_{\text{in}}(t)}{\Delta_{\text{out}}}, \quad (1)$$

provides a metric of the range of wind values during the calm [$\Delta_{\text{in}}(t)$] relative to the maximum variation between normal conditions and the minimum wind during the candidate calm period (Δ_{out}). A threshold value of 0.33 was heuristically selected for $R_{ra}(t)$ below which the zero wind values at the test site were accepted as calms. Higher ratios suggest that the corresponding values at the test site are unrealistically low in comparison with nearby sites. An explanatory example of this is provided in Fig. 2c, where the wind is shown for two selected candidate calms (blue lines) at sites 8400301 and 7026042, respectively, showing two opposite situations. The reference series (black lines) are shown with an indication of how the Δ_{out} and $\Delta_{\text{in}}(t)$ ranges are calculated. The red dots correspond to flagged values. While the whole candidate calm at site 8400301 is supported by the reference series, some of the values at site 7026042 are not. The reference series shows changes in the wind during the duration of the calm, indicating that the wind was likely not zero constantly. This approach allows for keeping values corresponding to genuine calms and flagging data that were likely different from zero.

2) ABNORMALLY HIGH VARIABILITY

These errors, typically a consequence of technical issues like loose wires or datalogger problems (Shafer et al. 2000), are in general less common than the erroneous low-variability records. A common method for detecting them is the so-called step check (Meek and Hatfield 1994; Hubbard et al. 2005), which compares the differences between sequential observations to a threshold value searching for steplike behavior (see Fig. 3a). For differences greater than the threshold, both values are regarded as erroneous. A somewhat more sophisticated approach is the blip test. This test looks for spikes and dips (Fig. 3a)—that is, successive increases and decreases in values (Fiebrich et al. 2010)—and unlike the step check, it is able to discern the faulty records from the good records. The thresholds for both tests can be either single values fixed for the whole time series (Meek and Hatfield 1994; Fiebrich et al. 2010), variable thresholds dependent on the month of the year (e.g., Vejen 2002; Dunn et al. 2016), or framed within the behavior of the day (e.g., DeGaetano 1997).

The method applied in this work uses a blip (or temporal) test complemented with a spatial check. The combination of both tests allows for identification of three different error typologies: spikes and dips, steps, and long episodes, schematically represented in Fig. 3a.

The blip test compares wind speed differences between valid consecutive observations with thresholds

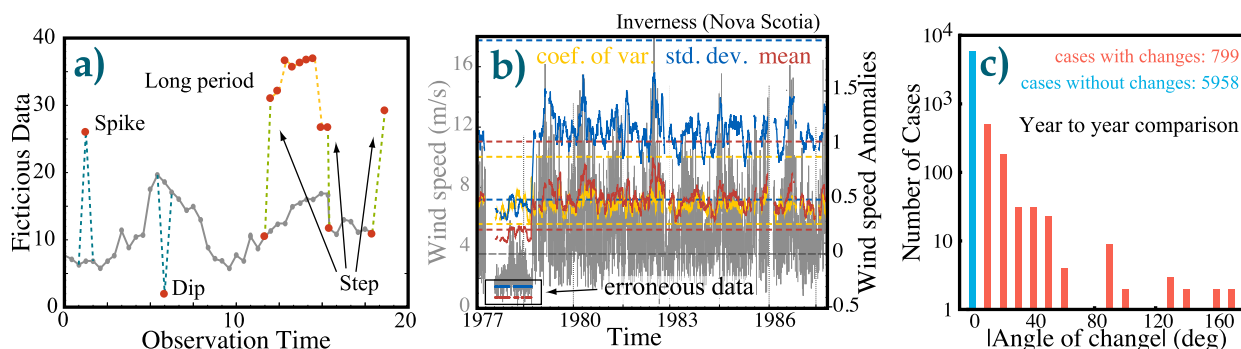


FIG. 3. (a) Conceptual illustration of high-variability errors. A normally behaving sample (solid line). The various types of detectable errors, namely, spikes, dips, steps, and long periods, are indicated as deviations (dashed lines; figure based on Fiebrich et al. 2010). (b) The wind speed bias detection method applied to station 8202550 (Inverness, Cape Breton, Nova Scotia; EC). The 15-day moving averages of the daily means of the anomalies (red line, right y axis), standard deviations (blue lines), and coefficients of variation (orange lines); horizontal lines of the same color depict their corresponding thresholds. The detected erroneous values of mean and standard deviation are highlighted, for easier visualization, by points of the same color (lower part of the plot). The original daily means (gray, left y axis). (c) Wind direction bias, the number of detected cases in the year-to-year analysis vs the angle of the shift (absolute value). The y axis is in logarithmic scale.

that are specifically defined for each station, so the analysis is run individually site by site. The differences that exceed the given thresholds are considered suspect. Since the series contain missing data and different time resolutions, consecutive observations may be separated by different time intervals. Therefore, different thresholds are obtained for each site from the distributions of the differences between pairs of observations separated by time intervals (Δt) ranging from 1 to 23 h. The time interval thresholds ($T_{\text{limit}\Delta t}$) are defined as $T_{\text{limit}\Delta t} = P_{95} + 8.9 \text{ IQR}$ from the distribution of differences. It was found that the positive differences were usually greater than the negative ones, meaning that the wind usually increases more abruptly than it decreases. The differences were subsequently split into negative and positive values, leading to $2 \times 23 = 46$ $T_{\text{limit}\Delta t}$ thresholds. At sites with low time resolutions, some of the short intervals are unlikely to happen and may lead to relatively small samples. Thus, for intervals with fewer than 100 cases, their corresponding thresholds were obtained from linear interpolation to the two closest $T_{\text{limit}\Delta t}$ values.

Additionally, each record is independently spatially evaluated using available wind speed data from the closest 10–40 sites located within a distance of less than 300 km and an elevation difference of less than 500 m (Dunn et al. 2016) to the target series. The record will be considered suspect if it exceeds the limit of its spatial threshold (S_{limit}), defined as $S_{\text{limit}} = P_{95} + 5.8 \text{ IQR}$. The combination of both tests allows us to flag unrealistic values within a temporal and a spatial context.

The suspect data may be classified into one of the aforementioned three categories (Fig. 3a). The blip test

by itself is able to detect erroneous spikes (dips) when a positive (negative) suspect difference is followed by a suspect negative (positive) one. In these cases, only the middle value is flagged. The erroneous steps are identified when a suspect positive (negative) difference involves also spatially suspect data. In those cases both values are flagged. Finally, all the values between flagged positive and/or negative steps will also be flagged if they fail the spatial test, constituting erroneous long periods. The validation of suspicious cases has been carried out by comparing them to auxiliary anemometers when possible, as in the case of DFO buoys (see section 2). For periods longer than a day, an additional 24-h window has been preemptively flagged.

b. Phase 5: Bias detection

The previous section targets erroneous periods of constant values or high-variability errors that are a few days long at most. However, longer intervals of time, such as weeks or months, that have systematic unusual values of mean and/or standard deviation will not be identified by the preceding analysis. The fifth phase of the QC (Fig. 1) deals with the detection of systematic errors (or biases) in both wind speed and direction. These errors, common to any meteorological variable, are related to a great variety of factors, such as changes in the measuring devices, different averaging methods, changes in anemometer heights, or changes in exposure or site relocation (e.g., Alexandersson 1986; Begert et al. 2003; Thomas et al. 2005; Wan et al. 2010). This work considers biases at different time scales. We correct for long-term wind speed biases from documented changes in anemometer heights. Regarding wind direction, we

target the detection and correction of biases, specifically shifts in direction, which may affect interannual to multidecadal time scales. We also look for errors caused by many other, often unknown, factors that affect the behavior of wind speed records for periods ranging from several weeks to months, longer than those targeted in the previous QC steps but shorter than long-term inhomogeneities.

1) WIND SPEED

As discussed in [Part I](#), although measuring heights should follow the international standard 10-m height convention ([WMO 1950, 1969, 1983, 2008](#)), in reality many sites may have suffered major changes in height through time with the evolution of measuring practices, thus inducing discontinuities in time series. If the exact measurement heights used at a given site through time are known, then there are different methods to standardize these records to a common height. The methods may range from simpler ones based on the wind power law ([Klink 1999; Pryor et al. 2009](#)) to relatively elaborated ones that account for atmospheric stability ([Thomas and Swail 2011](#)). In this work, the standardization to reference height z_{ref} (m) is done using the logarithmic wind profile ([Thomas et al. 2005; Wan et al. 2010](#)),

$$v_{z_{\text{ref}}} = v_z \frac{\ln(z_{\text{ref}}/z_0)}{\ln(z/z_0)}, \quad (2)$$

where v_z is the hourly data (m s^{-1}) at the measurement height z (m) and z_0 is the roughness length (m). For land stations, the roughness length has been derived from the USGS National Center for Earth Resources Observation and Science (EROS) Global Land Cover Characteristics Data Base (GLCCDB, V2p0; [Loveland et al. 2000](#)). The Weather Research and Forecasting (WRF) mesoscale model ([Skamarock et al. 2008](#)) has been used to relate the USGS static information to the desired z_0 values for summertime (Julian days 105–288) and wintertime (Julian days 288–105) seasons and for a $3 \text{ km} \times 3 \text{ km}$ gridpoint resolution. For buoys, the roughness length was chosen as $z_0 = 0.0016 \text{ m}$ ([Thomas et al. 2005](#)).

The measurement heights were compiled from several different metadata sources. The height information related to Canadian land stations was acquired mainly from a database that gathered information from many climate station inspection reports (SIRs) from Environment Canada's National Climate and Data Information Archives ([Wan and Wang 2006](#)). This information has been supplemented by looking through additional individual digitized SIRs obtained from EC.

For U.S. sites, the heights have been extracted from individual annual local climatological data (LCD) publication files obtained from the NCDC Image and Publication System (<http://www.ncdc.noaa.gov/IPS/lcd/lcd.html>). Finally, for the moored buoys, the information has been obtained from the Meteorological Service of Canada (MSC) buoy status reports archived by the DFO (<http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/index-eng.htm>; [Table 2; Part I](#)). The value of z_{ref} is different for each site and corresponds to the last known height that appears in the metadata.

The flagging of nondocumented wind speed biases is done using daily averages. The method screens anomalous behavior beyond the limits of the temporal consistency tests ([section 3a](#)). For this purpose, 15-day moving windows of the mean (\bar{v}), standard deviation (σ_v), and coefficient of variation ($\text{cv} = \sigma_v/\bar{v}$) of the daily time series are compared with an estimation of the typical range of the statistical parameters at each site. When the threshold is exceeded for any of these three parameters, the information about the date and parameter is stored. For the estimation of the usual ranges of variability, we use trimmed daily series where the values beyond the $\bar{v} \pm 2\sigma_v$ threshold are not considered to reduce the weight of the outliers. First, with the trimmed series an estimation of the mean annual cycle is obtained. This is calculated by averaging all the available values of each calendar day over all the years of available measurements at the site. The resulting 366-day estimate is filtered with a 15-day running mean, thus providing a smooth estimate of the annual cycle with daily resolution; missing days in a series that extend only a few years are interpolated. Second, both the original and the trimmed daily series are divided (normalized) by the annual cycle estimate obtained in the previous step, thereby diminishing the variability associated with the seasonal changes. Third, a 15-day moving window centered on each day is then used to calculate the running means, the standard deviations, and the coefficients of variation from the resulting original and trimmed normalized series. Finally, the smoothed 15-day filter outputs of the original series are screened for extreme behavior. This is done by comparing them to thresholds obtained from the normalized and then averaged trimmed series.

The upper and lower thresholds are defined using $\bar{v}_{\text{up}} = P_{95} + 3.7\text{IQR}$ and $\bar{v}_{\text{down}} = 0.5$, respectively, for the 15-day running mean of the wind speed series; $\sigma_{v_{\text{up}}} = P_{95} + 7.5\text{IQR}$ and $\sigma_{v_{\text{down}}} = 0.044$, respectively, for the running mean of the standard deviation; and $\text{cv}_{\text{up}} = P_{95} + 7.5\text{IQR}$ and $\text{cv}_{\text{down}} = 0.22$, respectively, for the running mean of the coefficient of variation. When the threshold is exceeded for one of the three

parameters during 15 days or more, the sequence is flagged. Shorter sequences of up to one week are flagged when more than one parameter threshold is exceeded. This comparison enables us to identify the most extreme intervals at weekly to monthly time scales. [Figure 3b](#) shows a practical example of the method.

2) WIND DIRECTION

The detection of biases in wind direction addresses the identification and correction, not the removal, of shifts in direction. The procedure considered herein is one of the few attempts reported in the literature ([Begert et al. 2003](#); [Petrovic 2006](#); [Gruber and Haimberger 2008](#)) and searches for temporal changes in wind roses. As in the case of biases in wind speed, the approach presented here is univariate, thus each series is treated individually and comparisons with regional references are avoided. The rationale for this is that wind direction at the surface is shaped by orography, thus complicating the identification of good reference neighbors within the spatial scales of intersite distances of our database. The method is based on the comparison of annual wind roses in search for relative shifts. The method is run individually for each series. In the first step, the method compares consecutive yearly wind roses. This is done by shifting them to find the relative angle in which their root-mean-square differences (RMSD) are at a minimum. This approach assumes that the distributions of wind direction at a site—that is, the wind roses—remain approximately stable through time with minor year-to-year variations and vary only slightly due to long-term changes in the atmospheric circulation. This is a realistic assumption as shown in [Fig. 3c](#), where about 90% of the year-to-year comparisons did not register any shift change and only around 4% involved shifts larger than 10° . Sudden shifts are presumed to be caused by changes in measurements, location, anemometer heights, surrounding environment, or artificial biases of any other nature. RMSDs are calculated for all the possible relative angles between the wind roses, depending on the precision with which wind direction is recorded (8, 16, 36, or 360 sectors). Precisions of 360 sectors are in practice reduced to 36 for this analysis in order to make results less noisy. Only years with at least 75% of availability of data are considered. This is still sufficient to avoid a large drop in yearly records while having robust estimates of the wind rose, that is, reducing the number of false positives related to subsampling. After a first check in which year-to-year steps are considered, only rotations larger than 10° are preserved for the second round.

The comparison of wind roses is expanded in a second step to the time intervals between the previously flagged years. The longer samples used to estimate wind roses in

this step make the results more reliable and allow for disregarding previously estimated changes that might have been due to data paucity. The resulting cases are individually inspected. Rotations in wind roses with equiprobable wind directions are discarded. The remaining cases are corrected by means of addition/subtraction of the rotated angle to match the position of the most recent time interval. The corrections are checked to be consistent with the surrounding orography where possible. The available metadata have also been consulted when looking for information that validates our findings and for problems that might be too short or involved fairly small angles to be detected. The analysis is limited to complete years defined from 1 January to 31 December, which means that there is a chance of erroneously modifying some correct months or overlooking erroneous loose months belonging to mostly correct years. This problem should, however, involve a relatively small number of months, as the needed amount of data to discernibly alter the wind rose is generally large and tends to be even larger for smaller shifts (not shown).

c. Phase 6: Isolated records

Once the previous steps of the QC are completed, a final step (phase 6 in [Fig. 1](#)) is conducted to flag isolated suspicious data. After the application of a number of tests, it is not uncommon to find short groups of isolated data between relatively long segments of missing or flagged data. This can also happen in the original data series between relatively long periods of missing observations. The reliability of these data is questionable ([Lawrimore et al. 2011](#)). The criterion followed herein is that any sequence of observations of 24 h long or shorter that is surrounded by an interval of missing observations of 24 h or longer has been flagged.

4. Results

This section reports on the results of the QC process and provides some additional technical details and examples of the identified error typologies. The number of the affected records in each phase is presented in [Table 3](#), where all the phases of [Part I](#) and [Part II](#) are sequentially shown, offering a general view.

a. Phase 4: Temporal consistency

1) ABNORMALLY LOW VARIABILITY

Low-variability checks address the detection of suspicious constant data sequences (see [section 3a](#)), making a distinction between noncalm (wind speed $\geq 1 \text{ m s}^{-1}$; direction $\neq 0^\circ$) and calm wind speed ($< 1 \text{ m s}^{-1}$) situations. A total of 9461 wind direction records were flagged for noncalm situations (0.02% of total direction data;

TABLE 3. Number of affected data during each phase of the QC (Fig. 1) for wind speed (WS) and wind direction (WD), and in total. Percentages, in parentheses, are given with reference to the initial number of WS/WD records (53 956 328 records each). Column 4 refers to both WS and WD as if the removal (Part I) or FR (Part II) had been applied, since the elimination of a WS or WD record implies the loss of the WS–WD pair. For FC data, column 4 is a simple sum of columns 2 and 3. Since some records may have been FR more than once, the total of the last row does not correspond to the sum of individual steps. Phases 1–3 synthesize the results in Table 5 of Part I. Also refer to Fig. 9.

Phase	Wind speed [Number (%)]	Wind direction [Number (%)]	Total [Number (%)]
1. Compilation*			
Chronological sorting	Ø*	Ø*	Ø*
Duplicated dates	2261*	2261*	4522*
Decoding errors	1 003 991*	Ø*	1 003 991*
Site displacements	1 078 337*	1 078 337*	2 156 674*
Measurement units	(EC + NCAR)*	EC*	99 580 581 (92.28)*
Observation time	EC*	EC*	93 237 056 (86.4)*
2. Duplication errors			
Intrasite	2818 (<0.01)	2453 (<0.01)	5640 (<0.01)
Intersite	2225 (<0.01)	69 432 (0.13)	138 864 (0.13)
3. Consistency in values			
Criteria (calms + north)*	—	539 510 (0.99)*	539 510 (0.45)*
Limits	288 (<0.01)	181 (<0.01)	938 (<0.01)
4. Temporal consistency			
Abnormally low variability (ALV)			
ALV ₁ ($\geq 1 \text{ m s}^{-1}$, $> 0.1^\circ$)	3498 (<0.01)	9461 (0.02)	24 562 (0.02)
ALV ₂ (<1 m s ⁻¹ , —)	190 933 (0.35)	—	381 866 (0.35)
Abnormally high variability (AHV)	2082 (<0.01)	—	4164 (<0.01)
5. Bias detection			
Height/rose shifts	8 563 779 (15.87)*	931 842 (1.73)*	9 461 199 (8.77)*
Undocumented biases	103 562 (0.19)	—	207 124 (0.19)
6. Isolated records	6906 (0.01)	6906 (0.01)	13 812 (0.01)
Total removed	272 681 (0.5)	88 433 (0.16)	696 634 (0.65)

* All the steps from phase 1 and the FC records and not taken into account to calculate the total of removed/FR data in the last row.

Table 3), affecting 54 sites in total (see Fig. 4a): 47 EC (with 8004 records) and 7 NCAR sites (1457). Fewer wind speed records, 3498 (<0.01%), were flagged with 33 sites affected (Fig. 4a): 32 in EC (3159 records) and 1 in NCAR (339). There is a remarkable absence of failures in buoys, where their longest periods are less than 6 h. These short lengths can be explained by their much higher measuring precision (1° and 0.1 m s^{-1}) but the lack of longer sequences is noteworthy. Many more low wind speed records were flagged, 190 933 (0.35%), affecting 468 sites (89% of the database; see Fig. 4b): 320 EC (93% of the dataset, 137 048 records), 21 DFO sites (52%, 19 529), and 127 NCAR sites (89%, 34 356).

The total number of handled constant sequences for tests 1 and 2 adds up to around 602 000 in direction (with a flagging rate of $\sim 0.02\%$), about 52 000 in speed (rate of $\sim 0.16\%$), and close to 7000 for calms (rate of $\sim 2.58\%$). The details regarding the number of flagged sequences and their associated false-positive ratios are summarized in Table 4. Given the small number of suspicious sequences, all the periods were individually

screened and the ones that looked unrealistic were flagged. Figure 4c shows two periods of simultaneous constant wind direction and speed data for a site located in Parry Sound, Ontario, Canada, and involving 5 days. Figure 2b shows the longest flagged calm identified by this method, which belongs to a buoy located in the Laurentian Fan and involves a continuous sequence of over a year. The longest calm periods belong to the raw buoy data that register near-zero wind speed values even if the anemometer had been destroyed in a storm. Such undocumented problems were not solved at the data compilation (see Part I, section 2).

The third test is based on spatial consistency and was applied only to calm situations. It accounts for the largest part of the flagged data: circa 32 000 from a total of 97 000 candidate calms showed at least one flagged record. Although the method targets calms of any length, the percentage of flagged sequences increases with length to more than 75% for sequences longer than 24 records (Fig. 4d, total calms in blue, flagged calms in red, and percentages in black). The majority of the flagged records within the intradaily sequences are

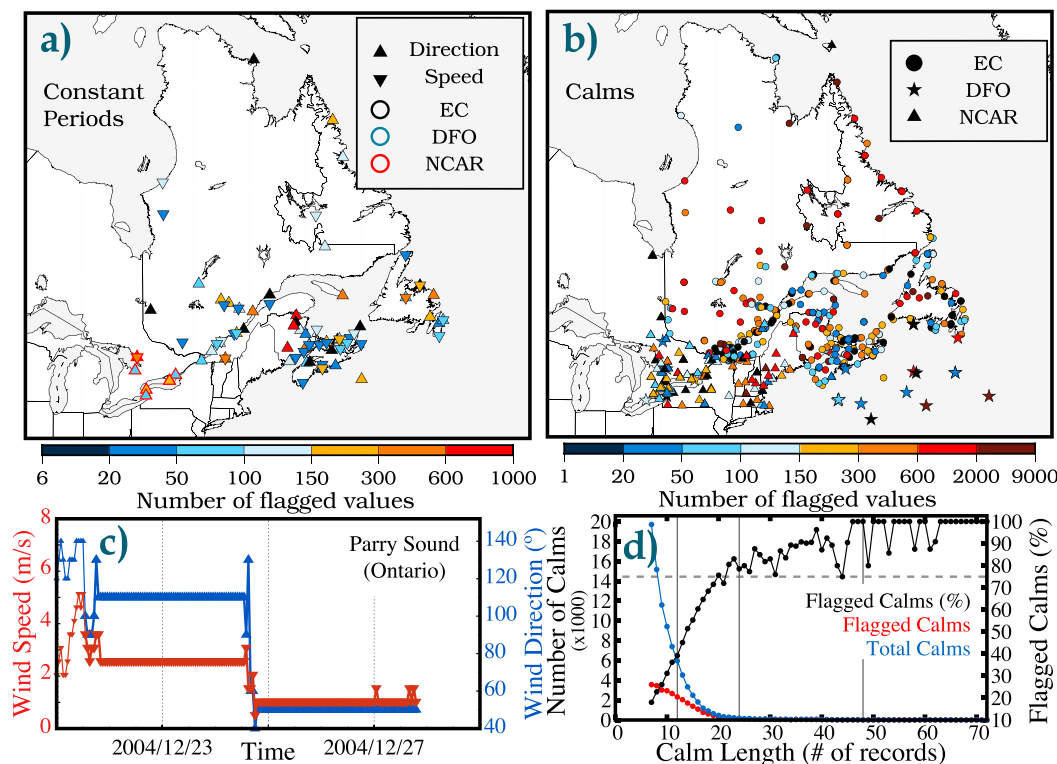


FIG. 4. Spatial distribution of the affected stations after checking for (a) long constant noncalm sequences and (b) calm corrections. Symbols indicate the variable (wind direction/speed) and source institution, and color fill indicates the number of affected data. (c) Example of simultaneous flagged constant periods for both wind speed (red) and direction (blue) at the same station (station CXPC, Parry Sound, Ontario; NCAR). (d) Absolute frequency distribution of the number of calms (left y axis) vs their length in number of records: those evaluated with a spatial comparison (blue), those that had at least one flagged record (red), and the percentage of flagged calms over the total evaluated (black, right y axis). In vertical (horizontal) bars lengths at 12, 24, and 48 records (percentages at 75%) are marked for easier visualization.

concentrated during diurnal hours when the winds tend to be higher and it is less likely that the regional series support zero wind speed at the test site (not shown). The percentage of flagged calm sequences remains more or less constant in the database for the whole time (not shown) with $\sim 10\%$ for EC and a little less for NCAR. DFO buoys show a disproportionate ratio that, at times, surpasses 50% of the cases.

2) ABNORMALLY HIGH VARIABILITY

Regarding high-variability errors, 2082 ($<0.01\%$) wind speed records were flagged, affecting 160 sites (Fig. 5a): 82 EC sites (528 records), three DFO buoys (3), and 75 NCAR sites (1551). The number of cases and associated false-positive ratios can be found in Table 4. The errors are more or less homogeneously distributed over time as shown in Fig. 5b and increase with the addition of new sites. NCAR sites show an abundance of flagged records, a consequence of the lack of QC processes applied to them. The majority

involves isolated records, placed well above the typical range of variability of the site, as shown in Fig. 5c for a site located in Massachusetts. Although less common, longer faulty periods can also be found in the database, such as the one in New Brunswick, Canada (EC, Fig. 5d).

b. Phase 5: Bias detection

1) WIND SPEED

The systematic biases in wind speed records are divided into two groups depending on their causes: those attributable to documented anemometer height changes; and those that, ranging from weeks to months, are caused by unknown/undocumented factors [see section 3b(1)].

The first group of biases was corrected following Eq. (2), which makes use of the available metadata on anemometer height changes. Documentation about the heights were available for 220 sites in total: 166

TABLE 4. Summary of the number of cases detected at each test of Part II. Results are provided in total and also separately for each data dataset/institution. The range of lengths of these cases is indicated in the number of records for abnormally low and high variability, and in hours/weeks/years for biases and isolated values. False-positive ratios, obtained after manual scrutiny of all the outcomes, are indicated within parentheses when possible.

Test	EC	DFO	NCAR	Total	Length range
Phase 4: Abnormally low variability					
Periods with excessive missing values ($WS \geq 1 \text{ m s}^{-1}$)	3 (—)	0 (—)	1 (—)	4 (—)	6–56
($WD \neq 0^\circ$)	20 (—)	0 (—)	1 (—)	21 (—)	6–54
($WS < 1 \text{ m s}^{-1}$)	57 (—)	4 (—)	35 (—)	96 (—)	6–96
Longest valid period ($WS \geq 1 \text{ m s}^{-1}$)	76 (23%)	0 (—)	6 (0%)	82 (22%)	16–143
($WD \neq 0^\circ$)	70 (23%)	0 (—)	12 (0%)	82 (19%)	25–508
($WS < 1 \text{ m s}^{-1}$)	63 (21%)	6 (0%)	16 (44%)	85 (23%)	21–8554
Spatial consistency ($WS < 1 \text{ m s}^{-1}$)	24 738 (—)	54 (—)	7254 (—)	32 046 (—)	6–1397
Phase 4: Abnormally high variability (WS)					
Temporal and spatial checks	247 (16%)	7 (57%)	1031 (0%)	1294 (3%)	1–40
Phase 5: Bias detection					
Documented height changes in WS	226 (—)	53 (—)	13 (—)	292 (—)	5 weeks–28 yr
Undocumented biases in WS	106 (19%)	16 (19%)	100 (12%)	222 (16%)	1 week–1 yr
Rotations in WD	59 (—)	9 (—)	27 (—)	95 (—)	1–17 yr
Phase 6: Isolated records					
Isolated values (WS and WD)	1161 (—)	291 (—)	4727 (—)	6906	<24 h

Canadian sites (EC + NCAR), all of the 40 DFO buoys, and 14 U.S. (NCAR) sites (Fig. 6a). Despite involving only 40% of the sites, these include 106 of the 125 sites longer than 20 years (85%, Fig. 6a), which are more prone to suffer from changes. The corrections have been applied to the 91 sites with at least one height change (Fig. 6b). The number of local documented changes may range between one and seven, with the longest sites suffering the most changes. Nevertheless, the comparatively shorter moored buoys can accumulate between one and four changes per site, due to a combination of two factors: 1) some buoys changed their hull type through time (Table 2 in Part I); and 2) each time series was constructed by combining the information of two channels (see section 2 in Part I), belonging to anemometers normally located at different heights in most of the cases (Table 2 in Part I). A total number of 8 563 779 (15.87%) records have been modified (Table 3). Figure 6c shows the temporal distribution of the documented heights for the 1953–2010 period. Measuring heights can broadly range from 37.19 to 6 m (10–3.3 m) for land sites (moored buoys). As we can see, before the late 1960s/early 1970s, there was not a preferred height as evidenced by the larger diversity of heights and the uniform distribution of stations over them (Fig. 6c). After the 1970s, however, a tendency to follow the standard 10-m height develops (Klink 1999; Wan et al. 2010) albeit with some notable exceptions (e.g., 37.19 m at the Greater Binghamton Airport, Binghamton, New York). The decrease in the percentage of 10-m-height sites after the late 1990s is parallel to the increase of heights below 10 m, which is mainly related to the appearance of moored

buoys. Figure 6d shows an example of one of the longest time series in the database (Goose Bay, Labrador, Canada; EC) with seven documented height changes. The records were corrected to its reference, last documented, height of 10 m.

Regarding nondocumented errors, 103 562 records (0.19%) were flagged (Table 3) with 78 sites affected (see Fig. 7a) in total: 37 EC (58 255 records), eight DFO (32 793 records), and 33 NCAR sites (12 514 records). Information about the number of detected cases and associated false-positive ratios can be found in Table 4. The most abundant flagged periods are under 4–5 weeks (Fig. 7b). Many of the shortest periods correspond to NCAR sites and coincide with high-variability errors detected previously. Two additional examples (Figs. 7c,d) show periods of about a month at Parry Sound (Ontario, Canada; NCAR) and about 3 months at Laterrière (Quebec, Canada; EC). The flagged cases from DFO tend to be of extreme low values, opposite of the cases of NCAR. The identified longest case (Fig. 3b) is likely due to an undocumented height change, as the coefficient of variation is unaffected (Vautard et al. 2010). The temporal distribution of flagged segments of data with biases (Fig. 7e) shows that despite the successive increase in the number of sites, EC has maintained a stable or even declining trend of erroneous periods in the last years, likely as a result of improvements in both site maintenance and data processing/QC methodologies. These numbers are comparable to those of DFO and NCAR in spite of the higher number of stations from EC.

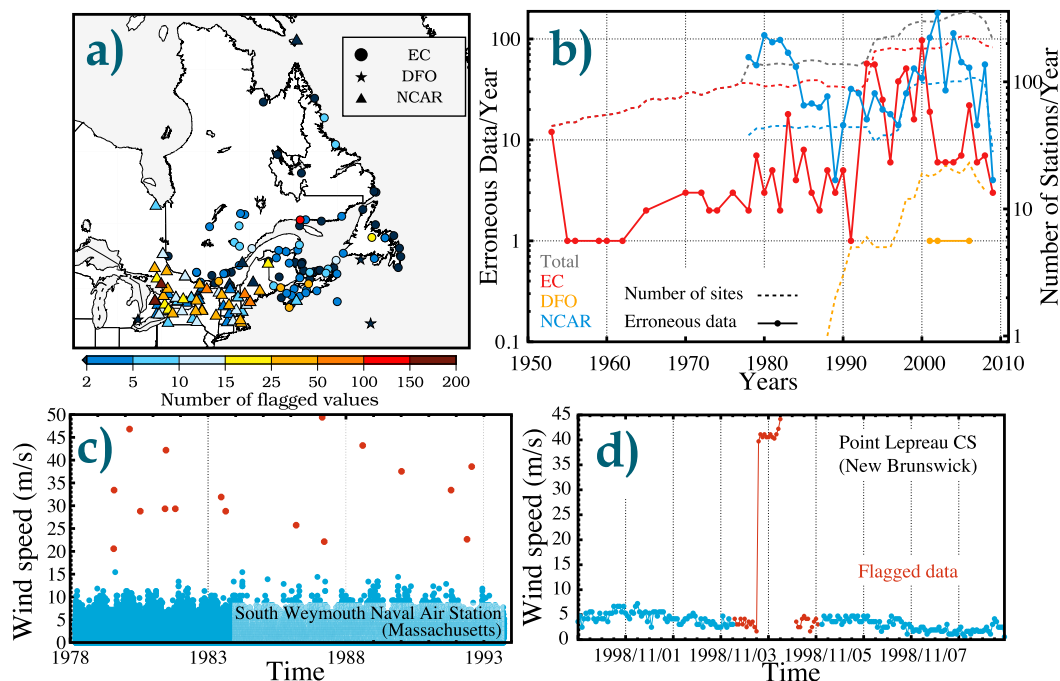


FIG. 5. (a) Spatial distribution of high-variability errors. Symbols indicate the dataset, and the color scale indicates the number of affected data. (b) Temporal distribution of flagged values. The number of operating stations per year (dashed lines, right ordinate axis), and the number of erased data per year (solid lines, left ordinate axis). Colors indicate the source institution. (c) Example of high-variability flagged data (red points) corresponding to site NZW (South Weymouth Naval Air Station, Massachusetts; NCAR). (d) As in (c), but for site 8104201 (Point Lepreau climatological station, New Brunswick; EC).

2) WIND DIRECTION

Detection of wind direction biases involved the correction of 931 842 (1.73%) records (Table 3). In total 36 stations were affected with vane shifts greater than 20° (Fig. 8a) with lengths expanding from one to several years. Most of these sites were affected by one or two shifts (circles). The metadata files (triangles) provided information to correct four additional sites, all of them with periods shorter than a year or with angles smaller than 20° and thus indiscernible by our method. In total 23 EC sites were affected with 670 798 records, three DFO buoys with 21 876, and 14 NCAR sites with 239 168 records. Figure 8b,c show two cases detected with our method: the first one corresponds to a site located at the Port Hastings Canal (Nova Scotia, Canada) with five changes and the second one corresponds to a station located at the Greenville Maine Forestry Service (Maine) with only one shift. An example of a shift detected from the metadata is shown in Fig. 8d. It corresponds to a site located in Blanc Sablon (Quebec) with a 10° angle shift that lasted from 20 March to 21 October 1974.

c. Phase 6: Isolated records

The last step in the QC process involved flagging 6906 (0.01%) pairs of records (Table 3) distributed among

6089 events of varying lengths of 24 h or shorter over 269 sites: 157 EC sites (2007 records), 22 DFO sites (291 records), and 90 NCAR sites (4608 records).

5. Impact

This section describes the impact of the whole QC procedure (Part I and Part II) on the statistics of the observational time series. Figures 9a,b show, for wind speed and direction, the type of error that had the largest implications at each site in terms of corrected/FC or deleted/FR data (excluding the phase related to compilation and the redefinition of *calms* and *true north*, in Part I). From an initial number of 526 stations, 501 have been affected with one or more of the analyzed nine error typologies in the case of wind speed. The most common error is related to unrealistic calms, being the most relevant one in 300 sites. The standardization of documented changes in height follows as the most important at 89 sites, followed by long-term errors at 52 sites and isolated values at 46. In wind direction, 310 stations were affected by any of the six analyzed error types. The most relevant error in the majority of the stations (209) is related to isolated values, followed by constant periods (42) and biases in direction measurements (40).

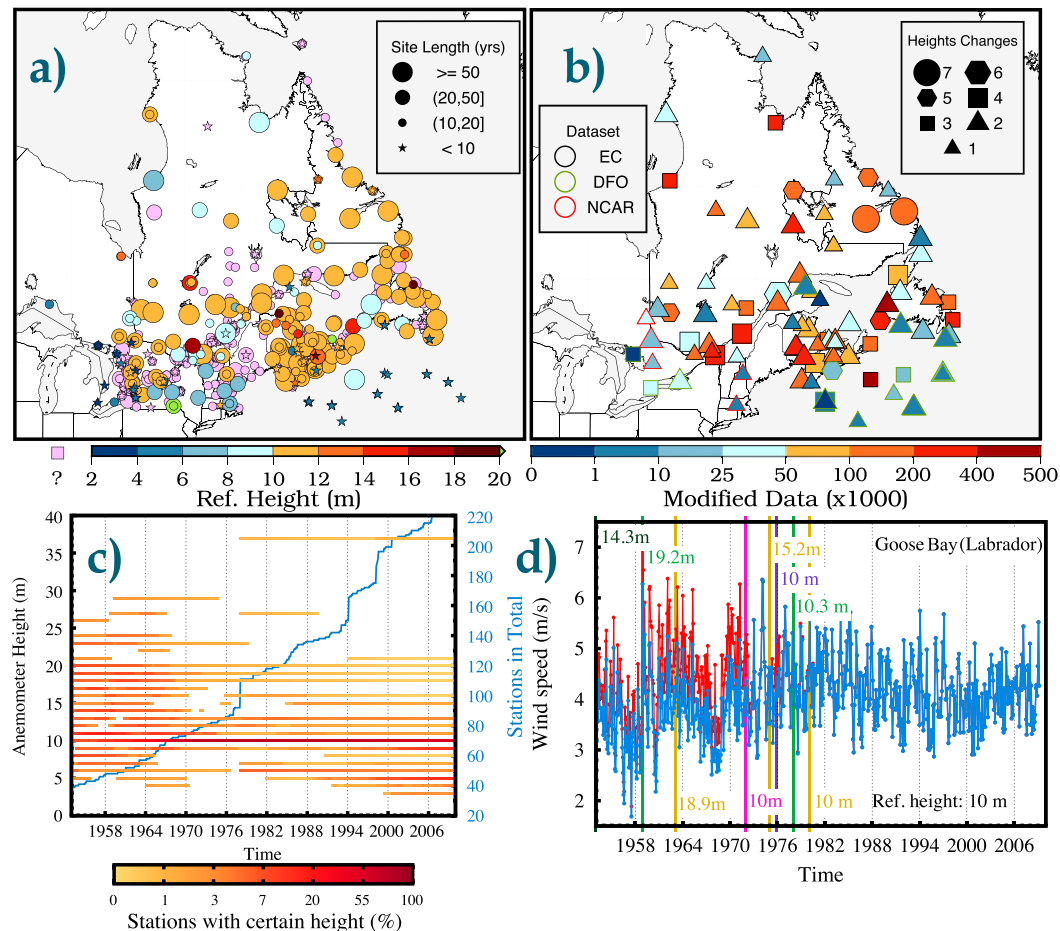


FIG. 6. (a) Spatial distribution of anemometer heights (m). The indicated height is the reference, last known, height. Sites for which individual information was not available (pink). The length of the time series is indicated with the size of symbols. (b) Spatial distribution of sites with at least one change in height. The modified amount of records is indicated by color filling, the number of documented height changes is indicated by the symbol, and the dataset is indicated by the color border. (c) Temporal distribution of known anemometer heights in the database. The heights (left axis) are rounded to the closest meter for clarity. The total number of active stations with documented height at each moment (blue line, right axis). The color bar shows the percentage of stations from the total that are placed at a given height (in logarithmic scale). (d) Monthly wind speed (m s^{-1}) of site 8501900 (Goose Bay) before (red) and after (blue) standardization to its reference height (10 m). Vertical color bars indicate documented anemometer heights.

Regarding wind speed and considering the different data datasets, those affected by one or more issues include 328 EC sites (95% of EC sites), 35 DFO buoys (87%), and 138 NCAR sites (98%). Regarding direction 183 EC sites were affected (53%), 27 DFO sites (67%) and 100 NCAR sites (69%).

The total accumulated percentage of removed/FR data by all the tests is shown in Fig. 9c. In total 501 sites were affected, although the vast majority of them (416) with less than 1% of data. Only six sites presented percentages above 10%, with four being buoys mostly affected by undocumented biases and very long calm situations that accounted for 25%–50% of their data

flagged. Site CWVY located in Lemieux (Quebec; NCAR; see Part I, section 2b) had all its data removed, as it was found that it was constructed with data of two other nearby sites. Fewer sites, 123, were affected by data corrections (accumulated percentages in Fig. 9d) but with higher percentages of affected data than in the previous case. Most of the sites (107) presented percentages above 10% and 33 sites more than 50%. Table 5 summarizes the number of FR/removed and FC/corrected data per dataset and in total.

Figure 9e categorizes the results by test and dataset, both with raw numbers and percentages. Although the EC dataset shows a higher number of flagged records

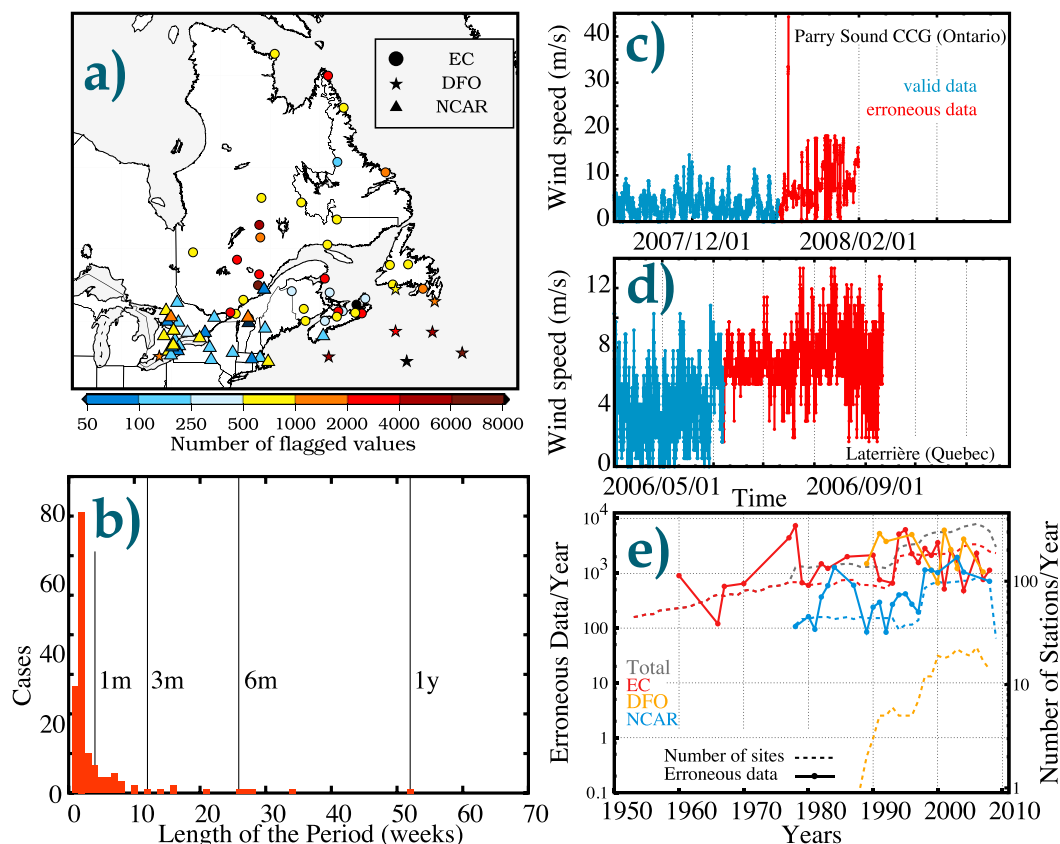


FIG. 7. (a) Spatial distribution of the sites affected by undocumented wind speed biases, symbols indicate the dataset, and colors indicate the number of affected data. (b) Number of cases vs their approximate length in weeks. Vertical bars indicate some time scales for better visualization. Examples of flagged periods of (c) almost 1 month at site CXPC [Parry Sound Canadian Coast Guard (CCG), Ontario; NCAR] and (d) 3 months at site 7064181 (Laterrière). (e) Temporal distribution of the flagged values. The number of operating stations per year (dashed lines, right y axis), and the number of flagged data per year (solid lines, left y axis). Colors indicate the source institution.

and had more sites affected than NCAR and DFO, when taking into account percentages the situation is reversed: NCAR tends to show more problems in isolated segments of data, unphysical measurements, problems related to the vane orientation, and with high variability; whereas DFO registers problems related with low-variability measurements and long-term biased periods. In percentages DFO (NCAR) has 8 (4) times as many removed/FR data as EC (Table 5).

The impact of the correction/removal of records on the shape parameters of the statistical distribution of data is shown in Figs. 10, 11. These parameters are the mean, standard deviation, skewness, and kurtosis obtained from the calculation of the first- to fourth-order moments (von Storch and Zwiers 2003). Despite not being the optimal estimators for non-Gaussian distributions, they have nevertheless been used, as they offer some valuable information about the changes in the wind distributions before and after applying the QC. The mean wind speed (direction) differences (before_{QC} – after_{QC}) are shown in

Figs. 10a,b). The majority of the sites were almost unaffected with changes smaller than $\pm 0.1 \text{ m s}^{-1}$ ($\pm 1^\circ$). Most of the speed changes are negative (the mean wind speeds after the QC are higher), mainly as a result of the removal of unrealistic calms. Some buoys on the east coast, with the longest erroneous calms, had the highest negative changes in mean. The highest positive changes in mean wind speed are related to high-variability errors and problems in the miscoding of missing values, specifically two sites with differences larger than 200 m s^{-1} (Part I, section 3c; consistency in values). The highest impacts in wind direction correspond to sites seriously affected by long-term biases (see section 4b) or to buoys recording erroneous calms. The standard deviation ratio (before/after, Figs. 10c,d) was close to 1 in most of the cases. The sites that presented the largest changes in wind speed variability were those with a high number of miscoded missing values. For wind direction, ratios are close to one except for a few land sites and buoys that were notoriously affected by long-term wind speed or direction biases (Fig. 10d).

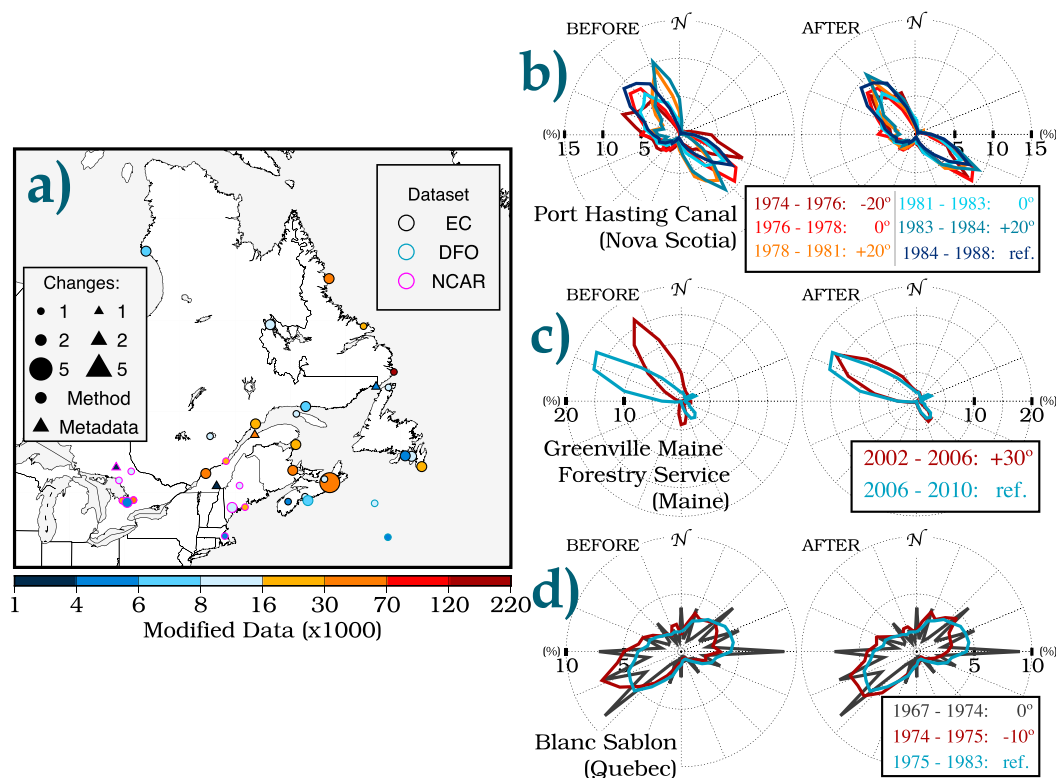


FIG. 8. (a) Spatial distribution of biases in wind direction. Colors of symbol contours indicate the dataset, symbols indicate that the shift has been detected with the method (circles) and metadata (triangles), symbol size indicates the number of vane changes, and the color scale indicates the number of affected data ($\times 1000$). (b) Wind rose showing wind direction bias before and after correction, corresponding to station 8204481 (Port Hastings Canal, Nova Scotia; EC) with five changes, each indicated by a line of a different color. The roses are shifted to match the last time interval that is considered the reference period. The indication for the rotated angle in the inset is provided also with respect to the last time interval. (c) As in (b), but for site KGNR (Greenville Maine Forestry Service; NCAR). (d) As in (b), but for site 7040812 (Blanc Sablon, Quebec; EC).

The skewness of the wind speed time series, a measure of the asymmetry of the distribution (Figs. 11a,b), is significantly reduced after the QC process, a sign of the effects of the unrealistic high records that have been erased on the tails of the distributions. Nevertheless, it continues being positive, which is a characteristic of this type of variable. It is noteworthy that all sites show now skewness values in a very close range of [0.25,1.75]. The kurtosis, a measure of the peakedness of the distribution (Figs. 11c,b), is also drastically reduced in stations with a greater number of high values but is still generally leptokurtic; here, the reference value of kurtosis is 0 for normal distributions. The stations now show a close range in kurtosis, $[-0.5,5]$.

6. Conclusions

This paper describes the second part of a semiautomatic QC procedure designed to identify and correct erroneous records of a surface wind speed and direction database of opportunity located in northeastern North America

(WNENA), compiled from an heterogeneous origin that were subjected to previous quality treatments of different depths. There are relatively few works covering these types of meteorological variables, especially at such depth (e.g., DeGaetano 1997; Graybeal et al. 2004; Dunn et al. 2016). The vast array of tests described herein provides an overview of data quality issues and offers some guidelines on how to improve a starting position that may not be optimal in terms of having data of the best desirable quality, but it may be nevertheless representative of what one can commonly access or acquire in the course of a time-bounded study. Most of these tests are either improved versions of previous studies or have been newly developed for this work. The tests described in Part II (Fig. 1) are focused on the detection of measurement errors: errors produced at the moment of the measurement and related to instrumental faulty performance or calibration, or sitting exposure. In contrast, the first three phases, described in Part I (Fig. 1, shaded), were centered on issues related to data management: problems originated at the compilation

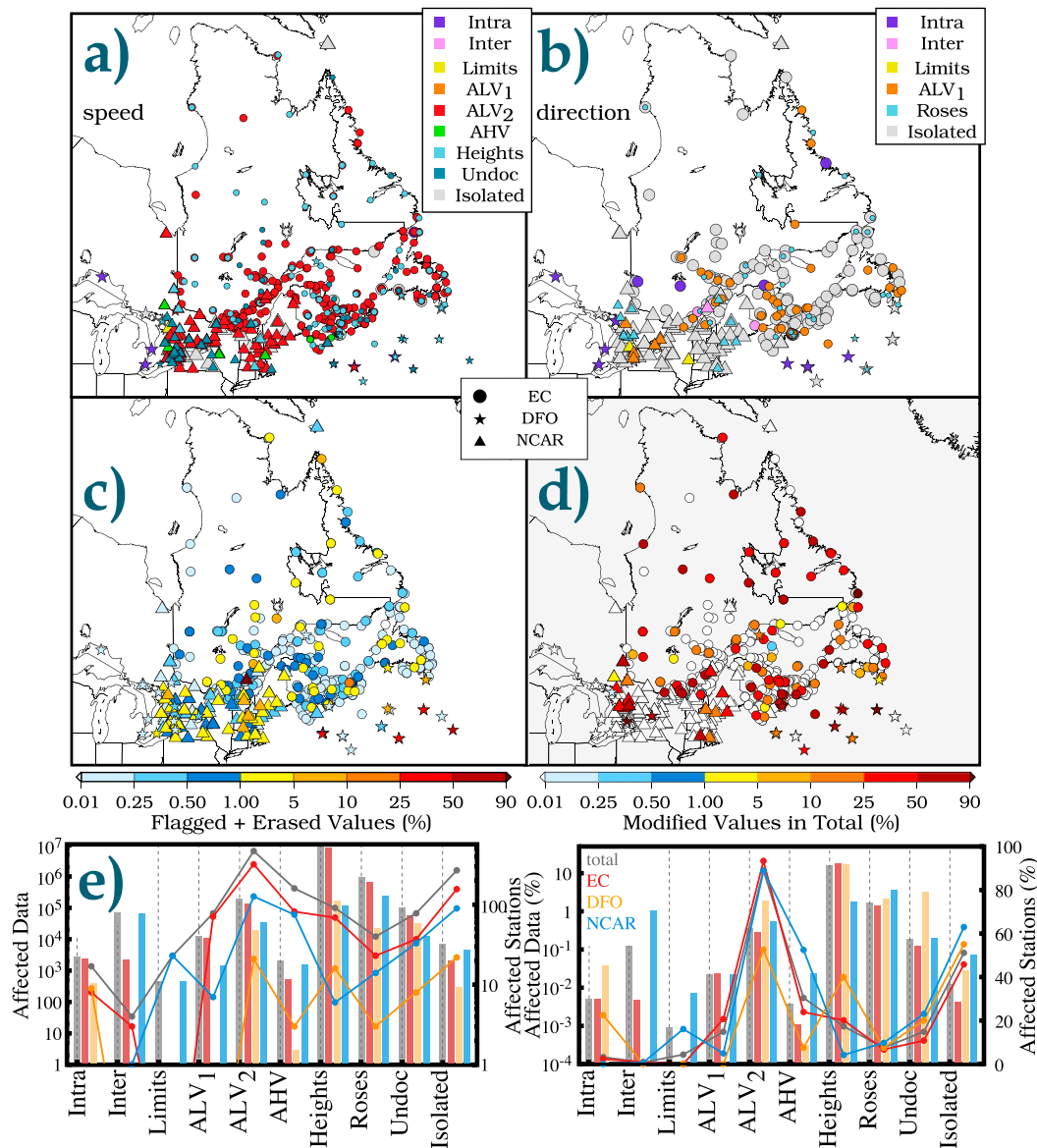


FIG. 9. Overview of the errors involving the largest amount of data at each site for (a) wind speed and (b) wind direction. Results for both Part I and Part II are included. The errors are indicated with colors, and the symbols indicate the data source institution (see legends). The symbols are given in decreasing sizes for easier visualization of close sites. (c) Distribution of the percentage of total deleted data at each site after all the steps of the QC. The stations with percentages over 90% were removed from the initial database. (d) Distribution of the percentage of total modified data at each site in Part II. (e), (left) Number of erased (Part I)/flagged (Part II) data (bars, left y axis) for the whole database (blue) and for each dataset (other colors), at each of the steps during the whole QC (x axis). The affected number of sites are in lines (right y axis). (f), (right) As in (left), but for percentages corresponding to total amount of data/sites (gray) or divided by dataset (rest). For the meaning of the abbreviations, refer to Table 3, phases 2–6.

and subsequent unification of databases from institutions that follow different criteria, and data manipulation procedures.

Most of the tests presented herein are based on simple principles and are computationally affordable, offering admissible false-positive ratios of ~3% in the best-case scenario (abnormally high

variability) and ~23% in the worst-case scenario (threshold-based test for noncalm constant speed situations). Given the volume of the database ($\sim 1.1 \times 10^8$ records), a special effort has been placed on a quasi-automatic design where the tests are nearly automatic but at the same time allows for a manual screening of dubious cases.

TABLE 5. Summary of total (wind speed + wind direction), and FR/removed and FC/corrected records by dataset and in total. The records that have been FR by multiple tests are counted only once. The percentages relative to the size of the dataset are in parentheses.

	EC	DFO	NCAR	Total
Total data	92 909 742	1 987 770	12 385 030	107 912 656
FR/removed	195 739 (0.41%)	34 342 (3.45%)	118 236 (1.86%)	348 317 (0.65%)
FC/corrected	8 920 933 (9.56%)	186 084 (9.36%)	354 122 (2.79%)	9 461 199 (8.77%)

As a result of the whole QC process (Part I and Part II), about 0.5% of wind speed and 0.16% of wind direction records have been identified as erroneous and removed/FR (0.49% and 0.03%, respectively, corresponding to Part II alone; see Table 3), resulting in a total of 0.65% of discarded data pairs. Additionally 15.87% of wind speed and 1.73% of wind direction records have been corrected after testing for biases (Part II) and more than 90% of the records were modified in

one way or another during the compilation (Part I). The results of the different procedures provide evidence of the inferior initial data quality of the NCAR and DFO datasets used in this study, with a large majority of high-variability-related errors at the NCAR sites and low-variability errors at the DFO sites. These sets present overall a larger percentage ($>8\times$ for DFO and $>4\times$ for NCAR) of erroneous records than EC. The largest impact of the QC on the mean and standard

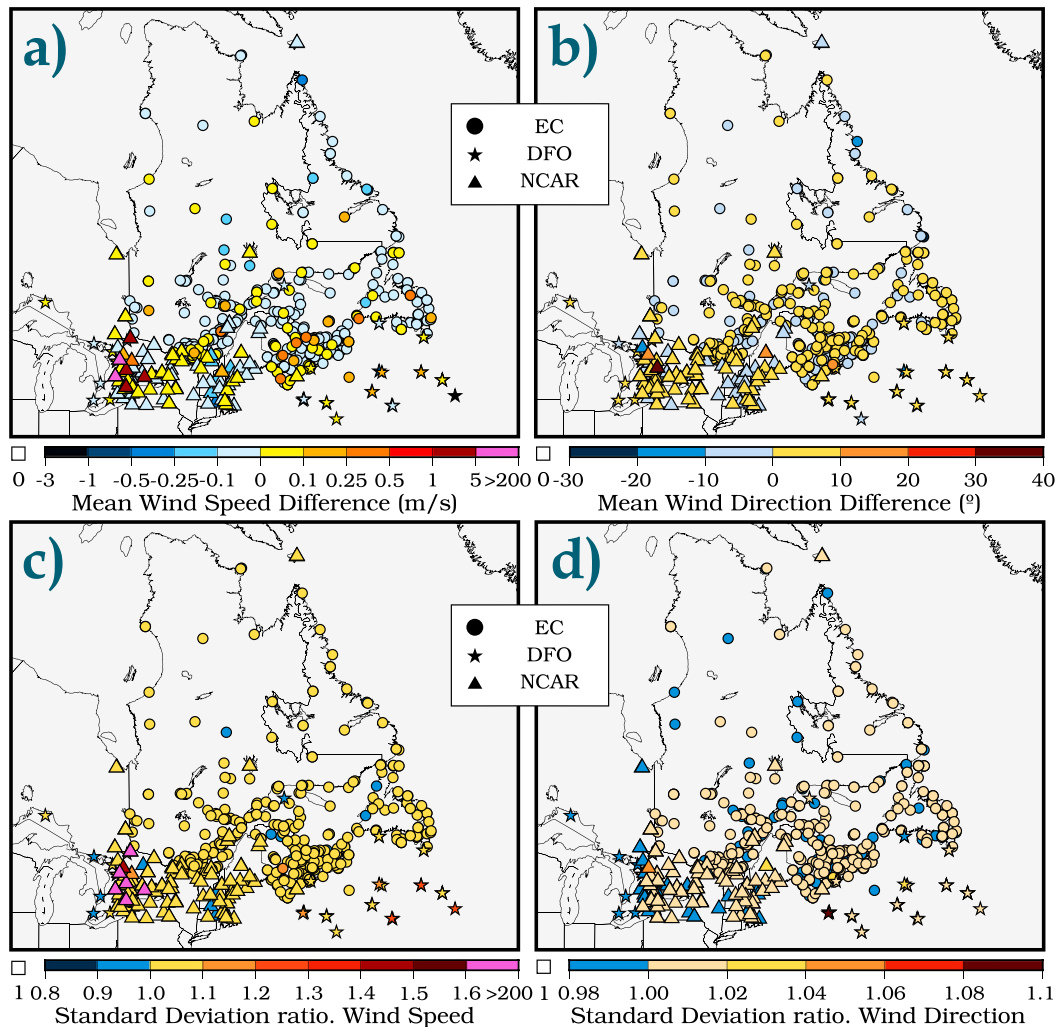


FIG. 10. Spatial distribution of differences in (a),(b) mean and (c),(d) standard deviation ratios for (a),(c) wind speed and (b),(d) wind direction (using directional statistics, Mardia and Jupp 2009) before and after QC changes.

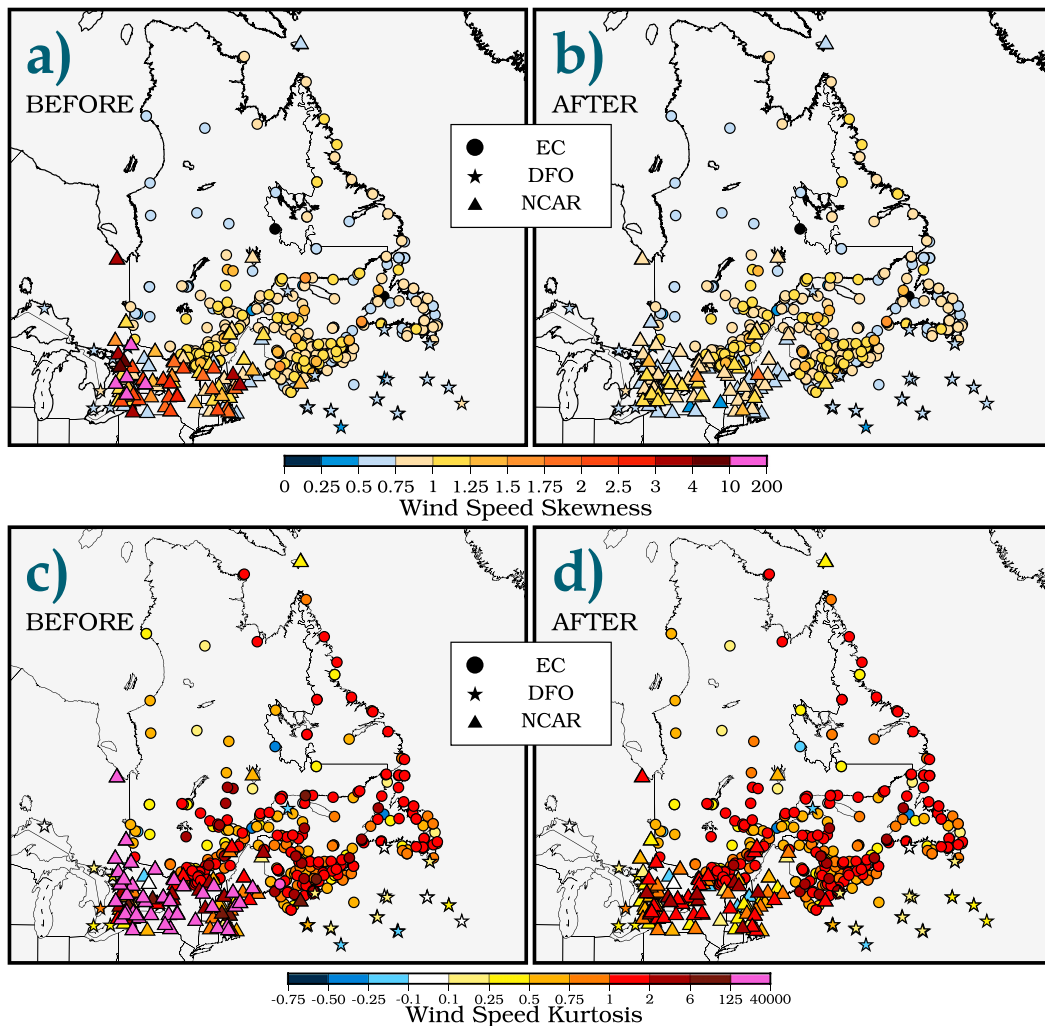


FIG. 11. Spatial distribution of the (a),(b) skewness and (c),(d) kurtosis for wind speed (a),(c) before and (b),(d) after the QC.

deviation of wind speed and wind direction distributions is associated with cases of miscoded missing values (only in NCAR), sites with long-term biases, and buoys recording erroneous calms. On the other hand, the largest impact on the shape and tails of the distributions (skewness and kurtosis) are related to sites with high-variability problems, which are also mostly NCAR sites. NCAR and DFO not only present more errors than EC sites, but their effects are also more noticeable. EC, although not free of errors, showed fewer incidences across the NCAR and DFO raw databases, both during the compilation and processing phases (Part I, sections 2 and 3a) and also during the application of the data quality procedures.

Some general considerations related to the assessment of measurement errors can be extracted from the development and application of the procedures described herein. Regarding low-variability problems, the length of purported calms or in general, of sequences

during which wind speed or direction are constant, can vary greatly depending on the resolution and instrument precision. The importance of segregating sequences according to resolution and precision on the distributions of extremes has been shown (Fig. 2a). For unrealistically long calm periods, threshold analysis allows for singling out obvious erroneous cases. Identifying erroneous shorter calm sequences with plausible lengths is a challenging task. The spatial comparison has been shown to be useful for these situations (Fig. 2c). The tests for high-variability errors have been successful in flagging erroneous clusters of data (Fig. 5d) and extreme events of sites regardless of their recording time resolution. It would be interesting to adapt a similar technique for wind direction in the future, which is hardly addressed in depth in the literature (DeGaetano 1997). The test employed to look for undocumented long-term errors in wind speed has been effective at bridging the

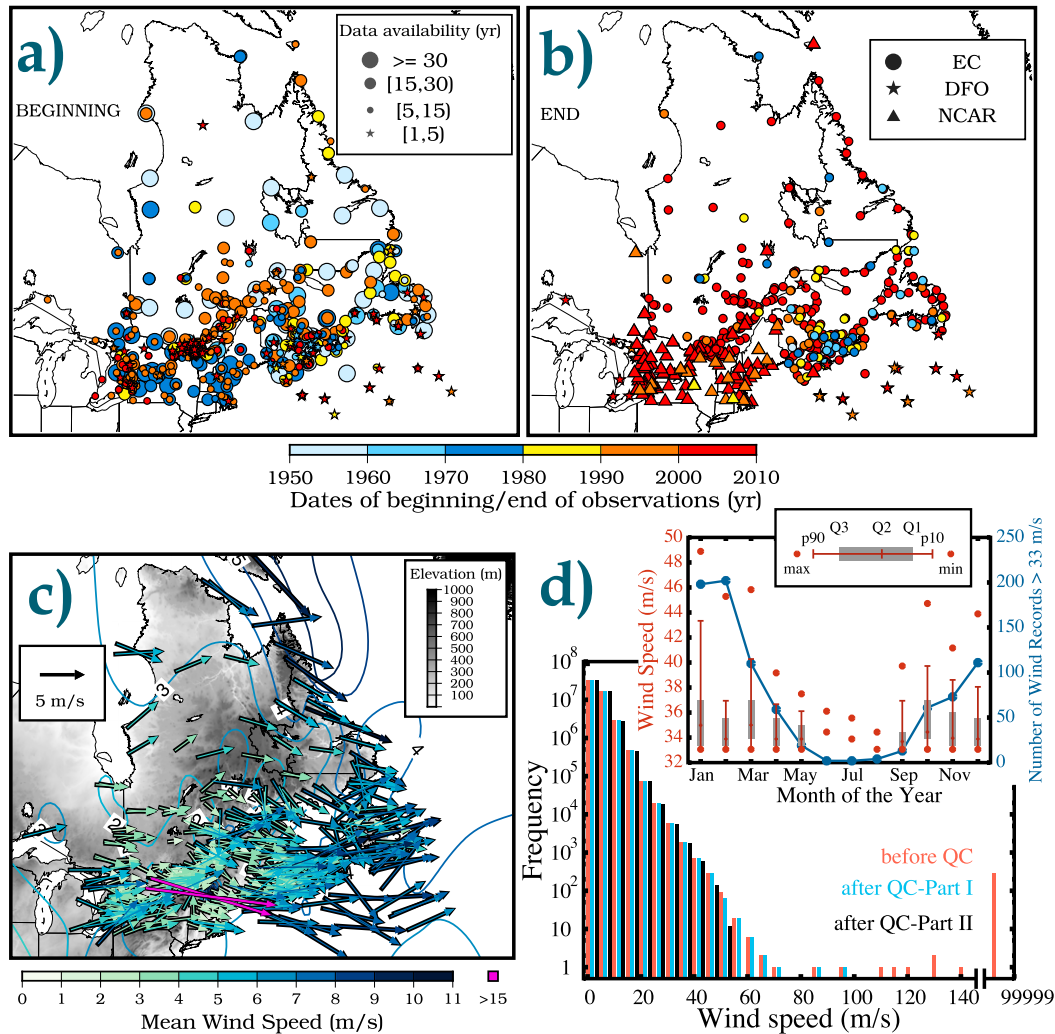


FIG. 12. Spatial distribution of availability of observations in the final database: (a) dates of first recordings at each site (colors) and the number of years with available data (symbol size), and (b) dates of last recordings (color scale). (c) Spatial distribution of mean winds (arrows) and standard deviations (isolines). The arrows give the direction from which the mean wind is blowing. The wind speed is given by the arrow size and color. The topography of the area is presented in grayscale. (d) Wind speed histogram comparing the pre-QC database (red) with the database after Part I (blue) and Part II (black). (inset) Seasonal distribution of the number of wind speed records $> 33 \text{ m/s}$ (blue, right y axis) after Part II; box-and-whisker plot of monthly distribution of wind speed records $> 33 \text{ m/s}$ (red, left y axis) also for the final database.

time-scale gap between the targets of traditional QC processes (hourly to weekly) and the statistical methodologies devised for homogenization problems (interannual and above; Fig. 7b). The use of metadata has been decisive in identifying a large number of changes in anemometer heights and their associated disturbances on long-term wind speed trends (Fig. 6d). Most of the corrections affected the longest sites, which are more prone to present successive height changes and buoys, with plentiful changes in hull types and transmission channels. Finally, a wind rose correction procedure (Figs. 8a,b) has been proposed with satisfactory results, a topic barely treated in the literature. The

method still poses some limitations regarding the minimum unit length for correction (one year) and the minimum detected angle of rotation (effective only for $\Delta\theta \geq \pm 20^\circ$).

The development and/or application of techniques to correct wind speed inhomogeneities of an undocumented nature, including long-term biases and/or drifts, has not been considered herein (e.g., Wan et al. 2010). Also, problems such as buoy tilt or wave sheltering in the DFO records (e.g., Gower 1996; Skey et al. 1998) are beyond the scope of the current work.

After the QC, WNENA consists of 525 sites. This database has a relatively homogeneous distribution of sites

through time (Fig. 12a). The oldest (and longest) stations are those starting in 1953. The database grows considerably after 1978 with the inclusion of some NCAR stations and also during the 1990s with the aggregation of DFO buoys and new EC and NCAR stations. The last 15 years of data show a spatially homogeneous and temporally stable coexistence of around 300 stations. A considerable number of stations (more than 200) are still active in 2010 (Fig. 12b), which would allow for expansion of the database in the future. Figure 12c shows the spatial distribution of the mean wind speeds and wind directions, and standard deviations of the database. The winds, predominantly westerlies, reach their maximum values along the coast of Labrador, the island of Newfoundland, and the Gulf of St. Lawrence. Figure 12d shows the effects of the QC on the wind speed distribution: before (red), after Part I (blue), and after Part II (black). As a result of the QC process, the highest realistic wind speed records have been reduced from 100 to 53.5 m s^{-1} . A box plot of the monthly distribution of hurricane-force-like records is presented in the inset. The majority of the events and highest wind values are distributed during winter, the season of the highest midlatitude storms' activity (Plante et al. 2015), and as with the higher mean values, they are also located along Labrador and the Gulf of St. Lawrence (not shown). Because of its spatial and temporal extension and resolution, WNENA is the database that, to our knowledge, best covers this region, offering a great opportunity to study the wind behavior from local to regional scales and from intradaily to multidecadal time scales.

Acknowledgments. EELE was supported by the Agreement of Cooperation 4164281 between the UCM and St. Francis Xavier University, and projects CGL2014-59644-R and PCIN-2014-017-C07-06 of the MINECO (Spain). Funding for 4164281 was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC DG 140576948), the Canada Research Chairs Program (CRC 230687), and the Atlantic Innovation Fund (AIF-ACOA). HB holds a Canada Research Chair in Climate Dynamics. JN and JFGR were supported by projects PCIN-2014-017-C07-03, PCIN-2014-017-C07-06, CGL2011-29677-C02-01, and CGL2011-29677-C02-02 of the MINECO (Spain). JC was supported by Global Forecasters until March 2014. This research has been conducted under the Joint Research Unit between UCM and CIEMAT, by the Collaboration Agreement 7158/2016. The research has also received funding from the European Union's Horizon 2020 Programme (2014–2020) and from Brazilian Ministry of Science, Technology and Innovation through Rede Nacional de Pesquisa (RNP) under the HPC4E Project (www.hpc4e.eu), Grant Agreement 689772. We wish to thank the people of Environment and Climate Change Canada, Department of Fisheries and

Oceans Canada, and the National Center for Atmospheric Research for providing us with the original data used in this study and for their kindness in responding to all the questions that arose during the development of this work and the review process. Special thanks to Gérard Morin and Hui Wan for the metadata of EC sites; Bruce Bradshaw, Mathieu Ouellet, and Bridget Thomas for information regarding moored buoys; and Douglas Schuster for information regarding the ds461.0 and ds464.0 datasets. We thank J. Álvarez-Solas, A. Hidalgo, and P.A. Jiménez for the helpful discussions. Finally, we would also like to thank the reviewers for the many suggestions and useful information they offered us.

Note: A first version of this database will be made available to the public. The QC procedures in this manuscript have been developed using Linux shell scripting and Fortran programming. Potential users interested in having the code are invited to contact the corresponding author.

REFERENCES

- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Int. J. Climatol.*, **6**, 661–675, <https://doi.org/10.1002/joc.3370060607>.
- Begert, M., G. Seiz, T. Schlegel, M. Musa, G. Baudraz, and M. Moesch, 2003: Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961–1990. *MeteoSchweiz Tech. Rep.* 67, 170 pp.
- Cheng, C. S., 2014: Evidence from the historical record to support projection of future wind regimes: An application to Canada. *Atmos.–Ocean*, **52**, 232–241, <https://doi.org/10.1080/07055900.2014.902803>.
- DeGaetano, A., 1997: A quality-control routine for hourly wind observations. *J. Atmos. Oceanic Technol.*, **14**, 308–317, [https://doi.org/10.1175/1520-0426\(1997\)014<0308:AQCRFH>2.0.CO;2](https://doi.org/10.1175/1520-0426(1997)014<0308:AQCRFH>2.0.CO;2).
- , 1998: Identification and implications of biases in U.S. surface wind observation, archival, and summarization methods. *Theor. Appl. Climatol.*, **60**, 151–162, <https://doi.org/10.1007/s007040050040>.
- Dunn, R. J. H., K. M. Willett, D. E. Parker, and L. Mitchell, 2016: Expanding HadISD: Quality-controlled, sub-daily station data from 1931. *Geosci. Instrum. Methods Data Syst.*, **5**, 473–491, <https://doi.org/10.5194/gi-5-473-2016>.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>.
- Fiebrich, C. A., C. R. Morgan, A. G. McCombs, P. K. Hall, and R. A. McPherson, 2010: Quality assurance procedures for mesoscale meteorological data. *J. Atmos. Oceanic Technol.*, **27**, 1565–1582, <https://doi.org/10.1175/2010JTECHA1433.1>.
- Gandin, L., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156, [https://doi.org/10.1175/1520-0493\(1988\)116<1137:CQCOMO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<1137:CQCOMO>2.0.CO;2).
- García-Bustamante, E., J. F. González-Rouco, J. Navarro, E. Xoplaki, P. A. Jiménez, and J. P. Montávez, 2012: North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula: Uncertainty and long term downscaled variability. *Climate Dyn.*, **38**, 141–160, <https://doi.org/10.1007/s00382-010-0969-x>.
- , and Coauthors, 2013: Relationship between wind power production and North Atlantic atmospheric circulation over

- the northeastern Iberian Peninsula. *Climate Dyn.*, **40**, 935–949, <https://doi.org/10.1007/s00382-012-1451-8>.
- Gower, J. F. R., 1996: Intercalibration of wave and wind data from TOPEX/POSEIDON and moored buoys off the west coast of Canada. *J. Geophys. Res.*, **101**, 3817–3829, <https://doi.org/10.1029/95JC03281>.
- Graybeal, D., A. DeGaetano, and K. Eggleston, 2004: Complex quality assurance of historical hourly surface airways meteorological data. *J. Atmos. Oceanic Technol.*, **21**, 1156–1169, [https://doi.org/10.1175/1520-0426\(2004\)021<1156:CQAOHH>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<1156:CQAOHH>2.0.CO;2).
- Gruber, C., and L. Haimberger, 2008: On the homogeneity of radiosonde wind time series. *Meteor. Z.*, **17**, 631, <https://doi.org/10.1127/0941-2948/2008/0298>.
- Hubbard, K., S. Goddard, W. Sorensen, N. Wells, and T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Oceanic Technol.*, **22**, 105–112, <https://doi.org/10.1175/JTECH-1657.1>.
- Jiménez, P., E. García-Bustamante, J. González-Rouco, F. Valero, J. Montávez, and J. Navarro, 2008: Surface wind regionalization in complex terrain. *J. Appl. Meteor. Climatol.*, **47**, 308–325, <https://doi.org/10.1175/2007JAMC1483.1>.
- , J. González-Rouco, E. García-Bustamante, J. Navarro, J. Montávez, J. de Arellano, J. Dudhia, and A. Muñoz-Roldan, 2010a: Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation. *J. Appl. Meteor. Climatol.*, **49**, 268–287, <https://doi.org/10.1175/2009JAMC2175.1>.
- , —, J. Navarro, J. Montávez, and E. García-Bustamante, 2010b: Quality assurance of surface wind observations from automated weather stations. *J. Atmos. Oceanic Technol.*, **27**, 1101–1122, <https://doi.org/10.1175/2010JTECHA1404.1>.
- Klink, K., 1999: Climatological mean and interannual variance of United States surface wind speed, direction and velocity. *Int. J. Climatol.*, **19**, 471–488, [https://doi.org/10.1002/\(SICI\)1097-0088\(199904\)19:5<471::AID-JOC367>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-0088(199904)19:5<471::AID-JOC367>3.0.CO;2-X).
- Lawrimore, J., M. Menne, B. Gleason, C. Williams, D. Wuertz, R. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res.*, **116**, D19121, <https://doi.org/10.1029/2011JD016187>.
- Loveland, T., B. Reed, J. Brown, D. Ohlen, Z. Zhu, L. Yang, and J. Merchant, 2000: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.*, **21**, 1303–1330, <https://doi.org/10.1080/014311600210191>.
- Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro, and H. Beltrami, 2017: Quality control of surface wind observations in north eastern North America. Part I: Data management issues. *J. Atmos. Oceanic Technol.*, <https://doi.org/10.1175/JTECH-D-16-0204.1>, in press.
- Mardia, K. V., and P. E. Jupp, 2009: *Directional Statistics*. Wiley Series in Probability and Statistics, Vol. 494, John Wiley & Sons, 456 pp., <https://doi.org/10.1002/9780470316979>.
- Meek, D., and J. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteorol.*, **69**, 85–109, [https://doi.org/10.1016/0168-1923\(94\)90083-3](https://doi.org/10.1016/0168-1923(94)90083-3).
- MSC, 2013: MANOBS: Manual of surface weather observations. 7th ed. Amendment 18, Meteorological Service of Canada Tech. Rep. En56-238/2-2012E-PDF, 488 pp.
- NCEP ADP OGSO, 1980: NCEP ADP operational global surface observations. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive. Subset: February 1975–February 2007, accessed 1 January 2010, <http://rda.ucar.edu/datasets/ds464.0/>.
- , 2004: NCEP ADP global surface observational weather data, continuing from October 1999. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 1 January 2010, <http://rda.ucar.edu/datasets/ds461.0/>.
- Petrovic, P., 2006: Detection of inhomogeneities in wind direction and speed data. *Proc. Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, WCDMP-71, Budapest, Hungary, WCDMP, 83–90.
- Plante, M., S.-W. Son, E. Atallah, J. Gyakum, and K. Grise, 2015: Extratropical cyclone climatology across eastern Canada. *Int. J. Climatol.*, **35**, 2759–2776, <https://doi.org/10.1002/joc.4170>.
- Pryor, S. C., and Coauthors, 2009: Wind speed trends over the contiguous United States. *J. Geophys. Res.*, **114**, D14105, doi:10.1029/2008JD011416.
- Shafer, M., C. Fiebrich, D. Arndt, S. Fredrickson, and T. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **17**, 474–494, [https://doi.org/10.1175/1520-0426\(2000\)017<0474:QAPITO>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<0474:QAPITO>2.0.CO;2).
- Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Rep. NCAR/TN-475+STR, 113 pp., <http://dx.doi.org/10.5065/D68S4MVH>.
- Skey, S. G. P., K. Berger-North, and V. R. Swail, 1998: Measurement of winds and waves from a NOMAD buoy in high sea states. Preprints, *Fifth Int. Workshop on Wave Hindcasting and Forecasting*, Melbourne, FL, Environment Canada, 163–175, <http://waveworkshop.org/5thWaves/C4.pdf>.
- Thomas, B. R., and V. R. Swail, 2011: Buoy wind inhomogeneities related to averaging method and anemometer type: Application to long time series. *Int. J. Climatol.*, **31**, 1040–1055, <https://doi.org/10.1002/joc.2339>.
- , E. Kent, and V. Swail, 2005: Methods to homogenize wind speeds from ships and buoys. *Int. J. Climatol.*, **25**, 979–995, <https://doi.org/10.1002/joc.1176>.
- Vautard, R., J. Cattiaux, P. Yiou, J.-N. Thépaut, and P. Ciais, 2010: Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nat. Geosci.*, **3**, 756–761, <https://doi.org/10.1038/ngeo979>.
- Vejen, F., Ed., 2002: Quality control of meteorological observations: Automatic methods used in the Nordic countries. Norwegian Meteorological Institute KLIMA Tech. Rep. 8/2002, 109 pp.
- von Storch, H., and R. W. Zwiers, 2003: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp., <https://doi.org/10.1017/CBO9780511612336>.
- Wade, C. G. N., 1987: A quality control program for surface meso-meteorological data. *J. Atmos. Oceanic Technol.*, **4**, 435–453, [https://doi.org/10.1175/1520-0426\(1987\)004<0435:AQCPFS>2.0.CO;2](https://doi.org/10.1175/1520-0426(1987)004<0435:AQCPFS>2.0.CO;2).
- Wan, H., and X. L. Wang, 2006: Canadian special metadata database for climate data homogenization. Environment Canada Internal Rep., 29 pp.
- , —, and V. R. Swail, 2010: Homogenization and trend analysis of Canadian near-surface wind speeds. *J. Climate*, **23**, 1209, <https://doi.org/10.1175/2009JCLI3200.1>.
- WMO, 1950: Provisional guide to international meteorological instrument and observing practice. World Meteorological Organization Tech. Rep. WMO-8, chapter 6.4, 68 pp.
- , 1969: Measurement of surface wind. Guide to meteorological instruments and methods of observation, 3rd ed. Secretariat of the World Meteorological Organization Tech. Rep. WMO-8, chapter 6.4, 10 pp.
- , 1983: Measurement of surface wind. Guide to meteorological instruments and methods of observation, 5th ed. Secretariat of the World Meteorological Organization Tech. Rep. WMO-8, chapter 6.6.2, 14 pp.
- , 2008: Guide to meteorological instruments and methods of observation. 7th ed. World Meteorological Organization Tech. Rep. WMO-8, 716 pp.